

A suitable case for treatment

The decision to reprimand a leading US medical school for failing adequately to monitor the protection of human research subjects is a salutary action. It also underlines the need to ensure that monitoring is properly funded.

Last week's four-day shut-down of research involving the use of human subjects at the prestigious Duke University Medical Center in Durham, North Carolina, has elicited predictable complaints in the research community (see page 190). Some argue that the move smacked of federal heavy-handedness, pointing out — correctly — that the lapses at Duke identified by the National Institutes of Health's Office for Protection from Research Risks (OPRR) were procedural. They argue that even if Duke's Institutional Review Board (IRB) — the panel responsible for ensuring that government-funded investigators adhere to strict ethics rules — kept inadequate records, sometimes lacked quorums, and allowed university grants and contracts officials to vote on whether proposed studies passed ethical muster, these are hardly crimes that justify bringing a multi-million-dollar research enterprise to a grinding halt.

That may well be true. But it is hardly cause for defending a review board that, if OPRR documents are accurate, had let itself slide into an extraordinary degree of sloppiness. The Duke IRB approved informed-consent forms that inaccurately rendered the risks and benefits of the experiments. It requested substantial changes in protocols, and then failed to follow them up. It kept minutes that were so sloppy that it couldn't be ascertained later who had been in attendance for what votes, why changes were required in certain studies, and whether or why informed consent was waived in others.

Of even deeper concern, the ethics panel regularly failed to document the special conditions needed for approval of research on children. All these problems, and others, were brought to the university's attention by the OPRR in December. But it took last week's enforced shut-down of research to prompt the kind of action from Duke

that should have been forthcoming in the first place.

Still, for all the embarrassment it caused in Durham and the worry it is now causing throughout the academic community — where everyone is wondering if they might be next on the federal hit list — the events at Duke should be seen in a positive light. First, the university is now showing a commitment to meaningful change. Second, institutions nationwide have been given fair warning: far better to put their IRBs in order now than for the serious injury or death of a human subject to bring down far more draconian federal constraints.

But all the motivation in the world won't solve two underlying problems: IRBs are overworked and underfunded, faced with an explosion of increasingly complex research as a result of rapidly expanding biomedical horizons, growing congressional largesse and increased private-sector funding of research. Medical centres and their funding agencies must give more resources to support them.

Furthermore, the OPRR itself — whose importance has been amply demonstrated by the state of affairs revealed at Duke — is far too small to cope with the task it faces. With a mere \$2.5-million budget and just one full-time human-subjects investigator, it is no use pretending that this office can monitor all human and animal research funded by the Department of Health and Human Services, as is its mandate.

Next month, a special panel is due to make recommendations on OPRR's institutional placement and authority to Harold Varmus, the National Institutes of Health director. It may recommend the OPRR's elevation within the health department. But whatever it says, not much will change unless the office, or its new incarnation, is given the financial muscle its job requires. □

Celebrations for some

A welcome boost to the refurbishment of Britain's research infrastructure must be supported by other measures.

Champagne corks were popping around Britain's universities last week as a select band of research groups celebrated their success in the first round of awards announced by the new Joint Infrastructure Initiative (see page 187). The initiative is jointly funded by the government (which has put in £300 million [US\$486 million], plus a further £100 million from the Higher Education Funding Council for England) and the Wellcome Trust (which has matched the government's contribution). It is intended to allow top-level researchers to refurbish laboratories and buy state-of-the-art equipment. Those lucky enough to be selected — 37 were chosen, 20 deferred and 124 rejected — were virtually unanimous in their enthusiasm for the extent to which the money will allow them to maintain world-class research activity.

But the scale of applications reveals the depth of the problem. The need for such a fund was first identified by a national commission chaired by (then) Sir Ron Dearing. Its report, published in July 1997, estimated the costs of renovating the UK research infrastructure as between £400 and £500 million. The new initiative shows this to have been a serious underestimate; the first round alone attracted bids exceeding £800 million, and requests for an equivalent amount are

expected in the three further funding rounds to come.

By all accounts, the administration of the initiative has gone relatively smoothly. Concerns that the central role of the Wellcome Trust, both in providing almost half the total funding and in keeping an eye on the overall distribution, would result in a bias towards the life sciences, have failed to materialize. Indeed, more than half of the awards were allocated to the physical sciences.

The decision to award grants strictly on scientific merit, while inevitably rewarding those (such as the universities of Oxford and Cambridge) with strong research traditions, has not been restricted to such elite institutions. And Treasury auditors will be reassured by the rigorous scrutiny successful applicants say their bids received.

Some concerns remain. One serious worry is that the money to be made available is far less than could be productively used. Another is the issue of academics' pay; in voting for strike action last week, the Association of University Teachers pointed out that this has fallen behind that of other sectors by 37 per cent over the past 18 years. As long as university research is seen as a badly paid profession, it will create barriers to the recruitment of the best and brightest. Shining new buildings and equipment are only part of the solution. □



French geneticists raise worries over use of new genome funds

[PARIS] Many French geneticists and some stock-market analysts are concerned that the government may be planning to prop up the biotechnology company Genset. They suspect that a FF1.5 billion (US\$240 million) boost to genome research, reported to be under consideration by the French government, may represent a hidden subsidy to the ailing French company.

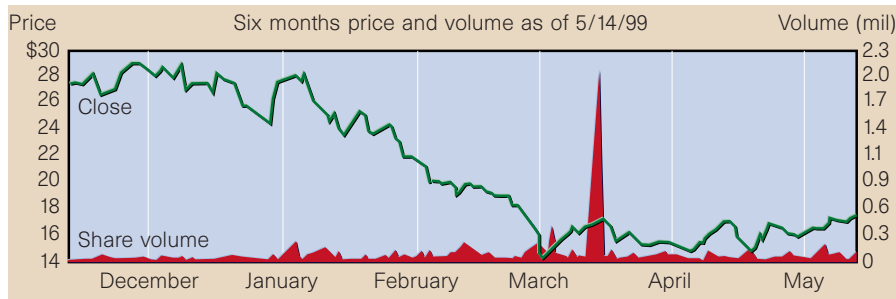
Shares in Genset, the largest biotechnology company in Europe and often called the 'jewel' of French biotechnology, have dropped sharply over the past twelve months from almost \$40 to around \$15. The company, valued last summer at over \$700 million, is now worth about half of this.

Genset's change in fortune has — according to several analysts — been largely prompted by market anticipation of the recent creation of a private-public sector consortium that would map and make freely available variations in the human genetic code linked to diseases. Genset's core business is to find and patent gene-sequence mutations involved in common diseases.

Pascal Brandys, chief executive officer of Genset, disputes this interpretation. He says the drop is "conjectural" and that other biotechnology companies are having the same experience.

He claims that Internet companies, for example, are drawing away investment from the sector and predicts that Genset's troubles are "provisional".

Whatever the explanation, information from a number of sources suggests that the company is likely to be the prime beneficiary of a possible FF1.5 billion injection of state



Falling fortunes: the popularity of Genset, the 'jewel' of French biotechnology, has dropped sharply.

funding into genome research.

The government is believed to be examining a proposal that would result in these funds being paid to a consortium made up of Genset, the Institute of Genetics and Molecular Biology near Strasbourg, the Génopôle biotechnology park being built at Evry, near Paris, and the Evry-based National Sequencing Centre (Genoscope) and National Genotyping Centre.

According to several sources, the proposal may give almost half the funds to Genset. Vincent Courtillot, director general of research at the science ministry, declines to confirm or deny this account.

"No public announcement has been made, particularly no numbers given, in any case by our ministry," says Courtillot. "Our prime minister has clearly said that France intends to hold its rank in genome-related innovation, through an alliance between public and private funds. But no other precise numbers are available yet."

Indeed, while France pioneered early genome efforts — its Génethon laboratory

published the first physical and genetic maps of the human genome in 1992 (see *Nature* 359; 794; 1992) — such research has until recently languished through lack of support from successive governments.

Pierre Chambon, head of the Strasbourg Institute of Genetics and Molecular Biology that was built in 1994 with the support of Bristol-Myers Squibb, confirms that the FF1.5 billion proposal is under discussion, but says that no decision has yet been taken. He denies, however, that the proposal is intended to indirectly subsidize Genset.

"If I take part in an operation, it is not to refloat Genset, it is because it is worthwhile," says Chambon, who is rumoured to be a candidate to head any consortium. "The essence of the project is outside Genset."

But many geneticists are concerned by what they claim is a lack of openness in the discussions surrounding the proposal, alleging that the decisions are being reached behind closed doors.

This has further fuelled suspicions that the proposal may partly represent a hidden subsidy to Genset.

"It is very worrying," says one leading geneticist. "Government money should not be used to compensate for the industrial risks taken by Genset — risks that have been punished by the market." This view is shared by several other prominent geneticists.

"It is obvious that, as the principle industrialist in this sector, we will take part in any initiative in this area," says Brandys.

He dismisses assertions that its involvement would amount to a subsidy, arguing that state support until now represents only a few per cent of the \$1 billion it has raised on the stock market. "One can't say we are big consumers of public aid," responds Brandys.

If the company did receive a substantial injection of cash through the proposed consortium, the state would also benefit in terms of contracts and work carried out, according to Brandys.

"There would be positive effects for

Big boost demanded for France's life sciences

[PARIS] An expert panel set up by the French government is likely to advise prime minister Lionel Jospin next month to create a multi-billion-dollar life sciences programme. It is expected to aim at linking scientists and industrialists in a coordinated national effort.

The proposal is also expected to call for the creation of a research agency dedicated to genome and molecular research related to therapeutics, along the lines of the French Atomic Energy Commission (CEA). The idea is to repeat in

life sciences the success of the CEA, which has helped France to become a world leader in nuclear energy technology.

The panel, called Research and Medicines, is expected to submit its proposals for a national strategy for the scientific and economic development of the life sciences to Jospin within the next two weeks.

Bernard Pau, a member of the panel, declined to confirm its specific recommendations before they were given to the government. He said: "We

need to decide on a very big investment in big biology. The life sciences will be one of the big industrial strategic activities of the next century".

Pau sees a need for France and Europe to put biology on the same footing as the French-led European aerospace programmes, such as the commercial Airbus project.

"I want to see the birth of a biotech Airbus," he says, adding that he is "optimistic" about the government's response to funding a major programme. **D.B.**

Génopôle and for society at large," he says.

Support for a role for Genset comes from Eric Molinie, an official at the French Muscular Dystrophy Association. The charity has invested more than FF1 billion in the Evry Génopôle and was the largest funder of French genome research before it shifted its focus in the mid-1990s to gene therapy.

The consortium proposal is "a good way to unite public and private genome research", says Molinie, arguing that there are few genome companies in Europe with the size and expertise of Genset. The consortium would allow France to recapture the lead it has lost in genomics, he says.

But Charles Auffray, a leading French genome researcher, says that — while there is a clear need to relaunch genome research in France — he prefers an integrated approach that would assemble a broad spectrum of public and private bodies within a comprehensive pre-competitive programme. Discussions on the way forward need to be more open, says Auffray.

The need for a broad approach is also supported by Bernard Pau, director of biotechnology policy at the Centre National de la Recherche Scientifique, and a member of a panel set up by the government to recommend a new national strategy for life sciences (see previous page).

Pau argues that genome research itself is increasingly seen as being less important than

post-genome research areas. This view seems to be shared by the markets.

Genomics is now considered to be a "low-price commodity", says Genghis Lloyd-Harris, director of equity research for the pharmaceutical and biotechnology sectors at the investment bank, Credit Suisse First Boston.

"The market is more interested in post-genome research, such as proteomics and DNA chip technology, which it sees as being more directly related to drug discovery," says Lloyd-Harris.

Such reasoning is one explanation for market dissatisfaction with Genset. A more immediate cause for the fall in its shares, suggests Ernie Knewitz, from Noonan-Russo Communications — Genset's public relations firm — is the recent decision by ten of the world's leading pharmaceutical companies and Britain's Wellcome Trust to fund a US\$45-million map of the variations in the human genetic code linked to common diseases (see *Nature* 398, 545; 1999).

By agreeing not to patent the map and to make it freely available, the group has implicitly challenged Genset's core business, namely mapping such single-nucleotide polymorphisms (SNPs) and striking alliances with pharmaceutical companies. One industry analyst says the new initiative turns SNP mapping into "pre-competitive research".

Knewitz points out that Michael King, an

analyst at Robertson Stevens, has reiterated the firm's "buy rating" on Genset, in the belief that the company has a large head start on the consortium and should complete its own proprietary map by the end of the year. "Investor concerns have been overplayed," says the text accompanying the rating. "The high level of interest in SNPs validates the Genset approach."

But one analyst expresses concern that Genset failed to live up to its "emphatic promises on milestones in terms of the number of SNPs mapped. The market was angry in feeling that it was misled."

Brandys vehemently denies this, arguing that numbers of SNPs or progress in mapping are irrelevant. The wider system for gene discovery is what counts, he says, including large-scale genotyping and analytical methods. "We have always been prudent in our claims," he asserts. "All that counts is our portfolio of patents. Our only milestone is gene discovery."

But the analyst adds that Genset pales besides rivals such as Celera, set up by genome pioneer Craig Venter, "which has patented thousands of genes and has three of its own products in clinical trials".

The market response to any new government support for Genset is unpredictable. "We are working on the assumption that Genset will get a whole lot of money," says one analyst.

Declan Butler

Increase in German science budget a boon for women and youth

[MUNICH] Germany's coalition government, which made a pre-election promise to double research spending within five years, has approved an 11 per cent rise in the 1999 science budget.

The increase brings the science budget to nearly DM15 billion (US\$8.2 billion). The new money will be primarily used to support research in biotechnology and information sciences (see *Nature* 398, 7; 1999). In biotechnology and molecular medicine, for example, an additional DM233 million will be made available in project money, and a new programme on the plant genome will be launched.

It will also be used to secure Germany's future scientific base by improving career prospects for young scientists and women. For example, it will pay for a new postgraduate programme, the Emmy Noether programme — named after a famous German female mathematician — launched this year by the Deutsche Forschungsgemeinschaft (DFG), Germany's university grant-making agency.

The programme supports young scientists, who must be under 30 when they apply, to complete two years of research abroad and then return to Germany to set up a university research group. The group's operating costs will be covered for three years. Normally, German scientists are well into their forties before



Noether: inspires a new programme.

they have the opportunity to direct their own research group.

The Emmy Noether programme has been allocated DM14 million for the current year, but this will rise to DM60 million a year as the number of researchers supported grows.

The extra money will also be used to launch a DM7.5 million programme of special measures to promote women's career prospects. These include special coaching programmes for women identified as talented during postdoctoral work.

These coaching programmes will be designed to "prepare women for competition for senior posts — particularly in strongly male-dominated areas like information science — before they disappear from science for ever," says a ministry spokeswoman.

But she points out that measures to promote equality of opportunity for women are not "ghettoized" in this new programme — all ministry programmes will be regularly audited to assess efforts to promote equality. "The spirit pervades the whole ministry," she says.

The same spirit will permeate most of

Germany's research organizations. For example, the DFG's Emmy Noether awards can be used in part to pay for child care. Recipients may choose to work part-time for an extended period to allow them to build a family.

Moreover, in plans drawn up between the ministry and the Helmholtz Society (the umbrella group of Germany's 16 large national research centres), the society has pledged to promote 100 women to top positions in its research centres in the next few years. The Research Centre Jülich has already started an active recruiting campaign (see *Nature* 398, 550; 1999).

Other measures include support of an International Women's University that will be established for 100 days during next summer's EXPO in Hanover. One thousand women from around the world will attend the university to teach and carry out research on seven themes: the city, work, migration, the body, intelligence, the information society, and water.

The ministry's new money also increases by DM200 million the federal government's contribution to university building and large equipment, which has been capped at DM1.8 billion for years. Germany's Science Council, the Wissenschaftsrat, has long complained that the building programme was dangerously underfunded.

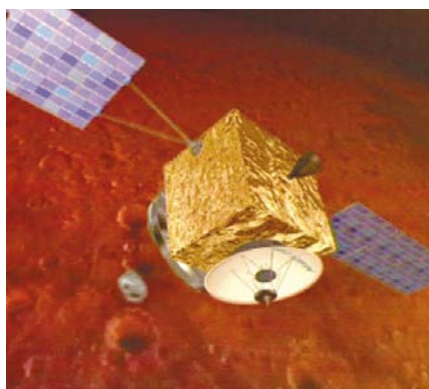
Alison Abbott

Europe keeps space science budget level

[MUNICH] Europe's space ministers last week agreed on a level budget for joint space science activities over the next few years that will allow science missions already approved by the European Space Agency (ESA) — including a mission to Mars known as Mars Express — to be launched on schedule.

The ministers also approved the start of a new optional programme of scientific missions to observe the Earth, called the Living Planet. Nearly 600 million euros (US\$640 million) has been approved for its first four years of operation.

But the meeting was protracted and acrimonious. Although 13 of ESA's 14 member states were prepared to agree an annual one per cent increase in cash terms to the mandatory space science programme for the next five years, Germany held out for a decrease



Full speed ahead: Mars Express is one of the missions to benefit from the budget agreement.

because it is already overcommitted in its own space budget (*Nature* 399, 4; 1999).

A hard-fought compromise limits the

financial envelope to four, rather than five, years — from 1999 to 2002 — and divides it into two components. The first is a 'baseline budget', a flat continuation of the current 352 million euro annual budget, which is not linked to inflation. The second component, averaging 10 million euros per year, is called 'complementary science funding'. Germany has committed itself to contributing only for the first year.

The science budget will be topped up through internal reorganization of resources, adding a further 9 million euros over the four-year period. This means that, despite the uncertainty of the complementary funding, purchasing power is likely to be maintained at 1998 levels.

ESA's director of science, Roger Bonnet, says ESA will therefore be able to launch Mars Express, so called because of the speed of its design and launch, as planned in 2003, and the infrared cornerstone mission FIRST — together with the microwave background mission Planck — in 2007. Other scheduled missions, such as the X-ray astronomy cornerstone mission XMM (due for launch in December 1999), will also keep to target.

Bonnet is due to discuss the scheduling of future missions with ESA's Science Programme Committee in Bern this week. The committee will also plan the selection of two new missions, each costing around 175 million euros. One will be in astronomy — possibly a contribution to the Next Generation Space Telescope, an international project led by NASA — and one will be in a non-astronomy field such as fundamental physics.

Unlike the mandatory science programme, ESA member states can choose the level of their involvement in the Living Planet programme. The programme has already attracted sufficient support among ESA members states to proceed with its plans to launch a small or medium-size scientific mission every year beginning in 2002 (see *Nature* 399, 4; 1999). This time Germany, where earth observation is a major priority, is the largest contributor.

The United Kingdom pledged a total of more than £100 million (US\$ 160 million) to four new space programmes. In announcing a £67 million investment in the Living Planet project, Lord David Sainsbury, Britain's minister for science and space, said the agreement marked "the first step towards providing an assured long-term programme of scientific research which looks at the Earth and its environment from space".

Other programmes to receive UK contributions are the ARTES 3 satellite multimedia programme (£27 million), the ARTES 1 satellite studies programme (£4.5 million) and the Galileosat navigation programme (also about £4.5 million).

Alison Abbott

Britain finds money to renew infrastructure

[LONDON] Christmas came early last week for 37 British research groups that are to share £150 million (US\$243 million) in the first round of a scheme to renew research infrastructure. Funding comes from the government and the Wellcome Trust.

In total, £700 million is to be spent on essential building, refurbishment and equipment projects under the Joint Infrastructure Fund (JIF). Researchers previously facing an uncertain future say they can now look forward to remaining internationally competitive. Against a backdrop of 20 years of underinvestment, a collective sigh of relief from the research community is almost palpable.

Individual awards are expected to range from one million to tens of millions of pounds (figures will be confirmed in the next few weeks). But most researchers admit to remaining nervous about the final sum they will receive. John Taylor, director-general of the UK Research Councils, says that the accounting for overheads and contingencies needs to be rigorously checked.

"We are clearly delighted," says Charles Godfray from Imperial College, London, a member of a team that put in a successful bid for £2.4 million towards new facilities for ecological research into climate change, biodiversity and microbial ecology. "This is absolutely tremendous. It is great seeing government invest in substantial infrastructure in universities."

At the University of Leeds, a project to develop, analyse and test transgenic crops will receive support. Three core programmes will look at transgenic

resistance to nematode parasites, signal transduction and control of gene expression, and whole-plant development.

A team from the University of Glasgow will receive funding to develop advanced detectors for the gravitational waves that are predicted to occur during violent astronomical events. Such detection will require highly sensitive equipment, and the team is working on advanced suspension systems for the mirrors within the detectors. Some £2.3 million will be spent on an ultra-high-vacuum room. Norna Robertson from the university says the money means they will remain internationally competitive.

At the University of Birmingham, a nanoscale science facility will receive support for what is claimed to be the world's first ultra-high-vacuum scanning transmission electron microscope. This will allow researchers to see below the surface of the three-dimensional nanometre-scale structures they build.

"We are learning how to fabricate on the nanometre scale," explains Richard Palmer of the university. The work will have an impact in areas such as materials science, chemistry, physics and electrical engineering.

The largest successful bid is thought to have come from a consortium of 18 UK universities which asked for £24.8 million for the Visible and Infrared Survey Telescope for Astronomy (VISTA). This will survey large areas for very faint stars and galaxies. It is intended to be "a fundamental source of reference on the contents and layout of our own Galaxy". The group are apparently "thrilled".

Natasha Loder

UK debates public's role in science advice

[LONDON] The British government has been presented with conflicting advice on the extent to which consumers and other public representatives should be involved in advising on regulations on the monitoring and control of genetically modified (GM) crops.

The government's ruling is planned to be published today (20 May) as part of a package of measures intended to calm public fears about such crops, while demonstrating that such concerns are being taken into account.

Last week, the House of Commons Environmental Audit Select Committee issued a report in which it proposed that members of the public should be appointed to the government bodies that are responsible for overseeing the safety of such crops. The committee argued that this would ensure that the scientific advice provided to ministers was tempered with ethical and consumer concerns.

On Tuesday of this week, however, the House of Commons Science and Technology Select Committee issued its own report on the role of science advice in regulating such crops. It argued that the make-up of scientific advisory committees should merely include suitably qualified experts from "other, not necessarily scientific, disciplines".

While also using the term "lay member" to describe such experts — whose role would



Hero or villain? Arpad Pusztai tells the select committee of his concerns over GM crops.

be "to ensure that the evidence is subjected to a sufficiently questioning review from a wide-ranging set of viewpoints" — the science select committee emphasizes that scientific advice should be offered free of any direct input from environmentalist or consumer representatives.

"Scientific advice must not be tinged with ethical and consumer concerns," says Michael Clark (Conservative, Rayleigh), the

chairman of the committee. "Scientists should concentrate on the science."

Such sentiments are reflected in the strong criticism the committee makes of press coverage of the potential impact of GM crops, including, for example, the prominent claims made by some newspapers about a link between increased antibiotic resistance — caused by the continued presence of antibiotic markers — and diseases such as meningitis (see *Nature* 397, 637; 1999).

To counter such reporting, the committee recommends that media coverage of scientific matters should be governed by a code of conduct "which stipulates that scientific stories should be factually accurate". It suggests that breaches of any such code of practice should be referred to the Press Complaints Commission.

The conclusions of both parliamentary reports were being studied this week by Jack Cunningham, the cabinet minister responsible for co-ordinating the government's response to the public outcry over GM crops, as he prepared to announce his decision on new regulatory arrangements today.

Prominent among these is expected to be some form of overarching body, bringing together the responsibilities of several existing advisory bodies, such as the Advisory Committee on Releases to the Environment, and the Advisory Committee on Novel Foods and Processes.

Also before Cunningham will be the results of a scientific inquiry carried out by a panel set up by the Royal Society into preliminary results revealed last year by Arpad Pusztai, a researcher at the Rowett Research Institute in Scotland. The panel found no evidence to confirm Pusztai's claim that his research indicated a potential threat contained in GM foods to the human immune system (see left).

Commenting on press coverage of the Pusztai affair, the Commons science committee — whose inquiry into GM foods forms part of a wider inquiry into the role of scientific advice in government — emphasizes the need to respect the crucial role played by the peer-review process.

"Those who start telling the media about alleged scientific results that have not first been thoroughly scrutinized and exposed to the scientific community serve only to mislead, with potentially very damaging consequences," says the committee's report.

The report also prompts the government "to establish international agreement on what constitutes a 'valid' scientific reason" and to ensure that "the definition of validity is based on the precautionary principle". This is an implicit reference to disputes such as that between Europe and the United States on labelling food produced using GM techniques (see *Nature* 398, 641; 1999). **David Dickson**

Royal Society: GM food hazard claim is 'flawed'

[LONDON] Britain's Royal Society announced this week that it has found no convincing evidence of adverse health effects from eating genetically modified (GM) potatoes in unpublished data from research at the Rowett Research Institute in Scotland.

The work was carried out by Arpad Pusztai, a senior researcher at the institute, who claimed to have detected a potential impact of such food on the immune system. This caused a nationwide controversy last year when it was described on a television programme (see *Nature* 394, 714; 1998 & 397, 547; 1999).

But a scientific panel set up by the society says that, on the basis of the information available to it, the work appears to be "flawed in many aspects of design, execution and analysis". The

panel adds that no conclusions should be drawn from it.

Pusztai told the panel that further data exists beyond that which was included in the review of his work. But he failed to produce such data or other evidence.

The panel's conclusions were published in a report on the Pusztai data and on toxicity in GM foods. The panel says that the slight differences that Pusztai claimed to have detected between rats fed predominantly on GM and non-GM potatoes could not be interpreted, because of the technical limitations of the experiment and incorrect use of statistical tests.

The panel also points out that, even if the experiment had been done skilfully, it would not be justifiable to draw conclusions about whether GM foods generally are harmful to human

beings. "Each GM food must be assessed individually," says the report.

Problems with the Pusztai data include a relative lack of information on how the GM and control diets differed in detailed composition. The GM potatoes used contained almost 20 per cent less protein than unmodified potatoes, and rats in the long-term feeding study were given additional protein to avoid starvation. The Royal Society panel suggests that the observed effects could have been caused by the supplementary diet being inadequate or incomplete. Its report says that Pusztai's work attempted to cover too much ground with the information available.

The society says its review of internal Rowett institute documents was "entirely appropriate" given that these are now in the public domain. **Natasha Loder**

University museum seeks independence

[SEATTLE] One of the larger and more unusual natural history collections in the United States is considering splitting from its parent university in a move that, while symptomatic of the low academic priority often placed on traditional museum activities, could have serious repercussions for teaching and research.

The Burke Museum of Natural History and Culture in Seattle is renowned for its ornithology specimens, including the world's largest collection of extended bird wings. It plans to become independent of its current owner, the University of Washington, largely because of a lack of financial support from the university to fund a modernization and expansion project.

Officials at Burke want to build a \$90 million facility to house specimens, create new research laboratories, and increase public outreach. But university administrators say this is such a low priority that it will not be funded. A Washington law limiting state expenditure has made matters worse, and Burke officials hope to be more successful at finding funding sources as a separate entity.

"In principle, the university has been very supportive. But in practice it hasn't been able to contribute the funds we need," says Karl L. Hutterer, Burke's director. "We came to the conclusion that it would be best for the Burke go on its own, maintaining linkage with the university through affiliation agreements."

This conclusion was reached despite the decision of a faculty-dominated committee in October that the museum should not separate from the university because of "serious, potentially very negative, consequences".

Washington provost Lee L. Huntsman says that, although it wants to maintain the "deeply treasured relationship" with the museum, the university is "feeling financially strapped" and needs to fund programmes closer to its "core mission".

Founded in 1885, the Burke museum has an annual operating budget of about \$2.6 million. It holds major collections involving mammalogy, anthropology, geology and zoology, including a genetic resources collection with unique frozen tissue specimens.

Following approval last month by the university's board of regents, Burke officials have set up a committee to determine how the museum could become a private, non-profit corporation, and to explore whether a new facility should be built on the university's main campus or at another location in Seattle. It will report by September to the regents, who must approve a separation plan.

Many systematics researchers argue that separating collections from a university has a deleterious impact on both teaching and research, hindering studies of biodiversity, ecology and comparative biology.



"Academic collections should remain with the university," says Peter B. Tirrell, associate director of the University of Oklahoma Museum of Natural History, and previous president of the Association of College and University Museums and Galleries. "There is a cumulative value of keeping research, teaching and students together."

According to Tirrell, Oklahoma conducted a similar assessment but chose to keep the museum at the university, assisted by a philanthropic drive that helped to build a new \$28 million museum. Hutterer points out that, although Seattle's economy has boomed with high-tech businesses, this has not been accompanied by such support for the Burke.

Many museums are now seeking to maximize the use of collections seen as low priorities by university executives, who have steered their institutions more towards molecular biology (see *Nature* 394, 115; 1998).

Earlier this month, the California Academy of Sciences and Stanford University in Palo Alto held a symposium highlighting the importance of systematics studies within museum and university communities.

One participant in favour of close ties between universities and museums is Brent Mishler, director of the University of California at Berkeley's University and Jepson Herbaria. "Those universities with the strongest programs in biodiversity are the ones with collections on campus," he says.

Ward Watt, an evolutionary biologist at Stanford, says he was "utterly disgusted" when Stanford moved its collection to the California Academy of Sciences in San Francisco 30 years ago. He says it meant that the university did not participate fully in the revolutionary changes in systematic biology.

As scientists probe plant and animal genomes, says Watt, collections take on more importance. He calls for studies by scientists from different disciplines using new tools.

Supporters of the Burke museum acknowledge the need for such collaborations, but argue that economic realities mean the museum has little choice but to try to fashion a new course.

Rex Dalton

More US labs may curb visits by foreigners

[WASHINGTON] Fermilab, Brookhaven and the US Department of Energy's other non-weapons laboratories could be caught up in the backlash against foreign visitors to government labs which has accompanied recent allegations of Chinese spying at the Los Alamos laboratory in New Mexico.

George Nethercutt (Republican, Washington) plans to amend an energy department funding bill to restrict visits of scientists from sensitive countries, including India, China and Russia, to all the department's laboratories, including civilian ones that hold no nuclear secrets.

The amendment would not ban such visits outright, but would require the Secretary of Energy to apply for special permission for visits involving a long list of prescribed technologies. The requirement would be lifted only after extensive new security and counter-intelligence measures had been implemented at the laboratories.

James Sensenbrenner (Republican, Wisconsin), the chair of the Science Committee in the House of Representatives, said he would support the Nethercutt amendment.

Laboratories hit by the proposed regulations would include Fermilab, Brookhaven, Oak Ridge, the Stanford Linear Accelerator Center, the Pacific Northwest laboratory and most of the other major physics facilities in the United States.

The Nethercutt amendment, which was due to be considered by the Science Committee this week, was watered down from earlier versions that would have blocked most visitors from sensitive countries altogether. But the level of support for the amendment suggests that the civilian laboratories can expect, at the very least, to implement tighter security in the wake of the Los Alamos allegations.

Scientific leaders were attempting to defend scientific collaboration with foreigners in the face of what a government official describes as strong public support for tough measures to clamp down on foreign visitors and to tighten security at US government laboratories. Both the National Academy of Sciences and the National Science Board are preparing statements to defend the foreign visitors programme at the laboratories.

Colin Macilwain

NIH ethics office clamps down on Duke . . .

[WASHINGTON] Duke University Medical Center (DUMC) has won a reprieve after the federal government shut down all its research involving human subjects last week. Federal regulators lifted the ban four days later, on condition that a number of remedial measures were taken immediately at the centre in Durham, North Carolina.

Among other things, the centre's Institutional Review Board (IRB), which is responsible for giving ethical approval to experiments, will have to re-review 274 of roughly 2,000 clinical trials which were approved when the panel lacked a quorum.

The centre will also have to immediately re-educate panel members and researchers on the federal regulations protecting human welfare in government-funded research. And it must set up a second IRB by 26 May, rather than by October, as had been planned, to relieve the overwork of the existing one.

Although there was no evidence that human research subjects have suffered at Duke, the decision on 10 May to shut down \$140 million-worth of government-funded research was taken because procedures to prevent this had not been properly followed.

The Office for Protection from Research Risks (OPRR) at the National Institutes of Health (NIH) cited 20 problems, ranging from meetings lacking quorums to conflicts of interest by voting IRB members who were university grants and contracts officials.

"The difficult work now begins," says Gary Ellis, director of the OPRR. "Building a broad and deep programme of protecting human subjects is going to take some time."

Ralph Snyderman, the chancellor for health affairs at DUMC, said in a statement



Under fire: Duke's medical centre

that officials there are "pleased" about the reinstatement. "We always have and will continue to place the safety of our study subjects as our highest priority," he said, adding that new programmes will "further ensure institutional oversight."

The shutdown is the third in seven months instigated by OPRR, the \$2.5 million NIH office responsible for ensuring the safety of human and animal subjects in research funded by the Department of Health and Human Services, of which NIH is a part.

Last autumn, OPRR shut down human-subjects research at a medical centre in Chicago for four days. In March, it did the same at the veterans' hospital system in Los Angeles (see *Nature* 398, 448; 1999), a move later extended by the Department of Veterans Affairs to encompass all research, basic and clinical, in the Los Angeles system (see box.) Before this, OPRR had not shut down human research since 1989, and it had never targeted an institution as well known as Duke.

Last week's action was "the nuclear bomb of enforcement," says Arthur Caplan, a bioethicist and former IRB member at the University of Pennsylvania. "At a major research university, to suspend funding even temporarily is an almost unprecedented event."

Some observers speculate that the pace of shutdowns and the high profile of Duke suggest that the OPRR is trying to toughen

up its image just as its political fortunes are coming into the balance.

In two weeks, the advisory committee to NIH director Harold Varmus is due to receive a report considering whether OPRR has enough authority and looking at its placement within NIH. Critics say it is hard pressed to execute its mission of regulating work sponsored by the agency which controls its fate (see *Nature* 397, 550; 1999).

The report is expected to recommend, at least as an option, that OPRR be elevated to the office of the secretary of Health and Human Services. "OPRR has ratcheted up the level of scrutiny and stringency [of IRBs]," says David Korn of the Association of American Medical Colleges. "One speculation is that OPRR has decided it's time to flex its muscles and show that it really is out there doing its job."

Others argued that even elevating OPRR within the Department of Health and Human Services would not give it enough independence. The suggestion that an independent body should monitor all government-sponsored research involving humans is hinted at in a letter to President Clinton made public by the National Bioethics Advisory Commission (NBAC) on 14 May, the day that Duke's research was reinstated.

Harold Shapiro, the NBAC chairman, writes that government ethics rules are difficult to enforce and improve "in part because no single authority or office oversees research protections across all government agencies and departments". The commission is due to make formal recommendations later this year for changes in the government system for protection of human subjects.

Ellis denies that his office has launched a get-tough strategy, and points out that the suspensions in Chicago and Los Angeles came after disputes lasting four and six years, respectively.

In none of the three cases did OPRR allege injury to human subjects. Rather, it targeted IRBs for lapses such as failing to include public representatives and failing to conduct continuing reviews of research trials that are required at least yearly.

The US media, advocacy groups and Congress however, have increasingly scrutinized human experimentation in the past two years. Several newspaper and television stories have focused on lapses in patient protection, particularly in psychiatric research.

This has also been examined by NBAC, which in December issued recommendations for improving research protection for the mentally ill. And the Inspector General of the Department of Health and Human Services last year issued a report saying that the IRB system was "in jeopardy" (see *Nature* 393, 610; 1998).

Meredith Wadman

. . . as basic scientists feel the pinch in LA

[WASHINGTON] Although the research shutdown at Duke University Medical Center (see above) only affected clinical researchers, basic scientists have been feeling the pain at the veterans' affairs (VA) hospital system in Los Angeles.

The NIH's Office for Protection from Research Risks shut down human-subjects research in the VA system in late March. But days later, a heavier blow landed: the Department of Veterans Affairs in Washington extended the ban to encompass research in the Los Angeles VA system in its entirety (see

Nature 398, 448; 1999).

The move was "equivalent to shutting down the US Air Force if you find a scandal in the Air Force Academy", complains George Sachs, a professor of medicine at UCLA and a senior medical investigator at the Veterans Affairs Medical Center West Los Angeles, where he is director of the laboratory of membrane biology.

Sachs, who studies the biology of duodenal ulcers, has lost the support of two NIH grants and a VA grant. He says he has put more than \$5,000 of debt on his credit card, buying everything

from DNA primers to computer software in order to keep experiments running.

Because local officials have said that "maintenance" research can proceed, Sachs is allowed to continue with some work that is already under way, but cannot start new projects. However, he says that many of his funding requests, which now must be routed through Washington, have been turned down. And some items are not to be had at all, including animals. So one postdoc, due to return to Germany this month, will leave without finishing his work on rabbit gastric glands.

M.W.

Chinese reform pushes R&D into market

[BEIJING] In a major reform of its public sector research system, the Chinese government has decided to convert into self-supporting enterprises 242 research institutes that now operate under ten government agencies.

The institutes "will go to market," says an official from the Ministry of Science and Technology. He adds that the government's goal is to push the institutes closer to industry, in the belief that this will increase the flexibility and vigour of public research and development (R&D), and speed up the application of research results in industry.

The institutes — with a payroll of 173,000 employees, including nearly 47,000 retired — will be free to choose their research and commercial strategies, according to a paper from the science ministry. But they will be strongly encouraged to become technological enterprises that can carry out R&D in addition to business activities.

The agencies include the bureaux of the coal, mechanics, metallurgical, textile and petroleum/chemical industries, as well as the bureaux of light industry and internal trade. Many were formerly ministry-level government agencies. In future, they will fall under the State Economic and Trade Commission, a powerful agency which reports to the State Council, China's central government.

The State Council has approved the reform plan, submitted by the science ministry and 11 other ministries. The institutes will be expected to operate on the new basis from 1 July. During a transition period, the State Economic and Trade Commission and the ten national administration bureaux will be responsible for routine management.

The science ministry spokesman denies that the government is taking this action because the institutes run inefficiently. Apart from a few, he says, most of the institutes are efficient and productive. "They have long learned to survive in the market," he says. "Indeed, 90 per cent of their revenues come from the market. So I don't think they will be greatly influenced."

But Sun Chuanyao, president of Beijing



Research in action: visitors at Chinese computer exhibition consider the latest equipment.

General Research Institute of Mining and Metallurgy (BGRIMM), is worried.

"After this reform, no one may be prepared to do work which is less lucrative, but which the state really needs," says Sun. "Certain research activities will have to be cancelled, and some institutes may not be able to survive as research organizations."

On becoming enterprises, says Sun, the institutes may find it hard to maintain a balance between earning money and carrying out fundamental research. "China is not short of enterprises, but most of the state-owned enterprises are debt-ridden. So-called 'technological enterprises' should avoid taking the path that they followed."

But he does not think the reform threatens his own institute. "We have been market-based for a long time, and have been in close relations with industry; this is our difference from the Chinese Academy of Sciences."

BGRIMM has a strong record. Its slurry electrolysis technology and its emulsified explosives for mining are world famous.

In 1997, the General Research Institute for Non-ferrous Metals (GRINM) drew a

400-millimetre long mono-crystal silicon line, 300 millimetres in diameter. The United States, Japan and Germany are the only other countries to have mastered this technique. In collaboration with other institutes, GRINM also produced China's first high-temperature superconductor cable made of materials from the bismuth series.

The government now faces the task of ensuring that institutes such as BGRIMM and GRINM remain creative and productive. "The government hopes that industrial enterprises will play a major role in China's R&D efforts, but such enterprises are far from that goal," says a BGRIMM researcher.

Although a few research institutes will be allowed to keep their government-supported status, they will have to operate like private companies. Administratively, they will be responsible to local government.

The research institutes involved in this reform will enjoy special privileges. They will continue to receive some operating funds from the government, although this is intended primarily to support those who have already retired. For the first five years they will be exempt from enterprise income tax, operating tax based on revenues from transferring technology, and property taxes on urban land used for R&D purposes.

The new bodies will have the same rights as remaining government research institutions when bidding for government-financed research programmes. Research programmes already approved by the government will be conducted according to the original arrangements.

"This is only the first step," says the ministry official. "The second step will begin immediately after this, when 400 or more research institutes under different ministries will be involved."

Tian Xuewen

Japan's fast-breeder loses money fast too

[TOKYO] Japan's Nuclear Fuel Cycle Development Organization (JNC) has built up a consolidated deficit of more than ¥1.6 trillion (US\$13 billion), says a finance report released last week by the government's Management and Coordination Agency.

The report says that JNC's losses are largely due to the high cost of research and development work at Monju, the experimental fast-breeder reactor. The programme has faced increasing public opposition after an accident at Monju in

December 1995 (see *Nature* 387, 646; 1995 & 389, 653; 1997).

JNC, formerly the Power Nuclear Fuel Development Corporation, has received ¥2.4 trillion from the government since 1996. Yet the value of its assets — mainly land and property — is currently only ¥850 billion.

The Management and Coordination Agency is now warning that JNC's research needs to be closely evaluated if it is to justify any future funding for the fast-breeder programme.

Asako Saegusa

Congress in move to block open access to research data

[WASHINGTON] Opponents of a measure that would open up researchers' data to requests filed under the US Freedom of Information Act are trying to amend this year's appropriations bills in Congress to halt the implementation of the measure for a year.

David Price (Democrat, North Carolina) and Jim Walsh (Republican, New York) were planning to put the amendment to the Treasury, Postal Service and Government appropriations bill for fiscal year 2000 when it passes through the full House Appropriations Committee this week. The amendment would block the White House Office of Management and Budget from implementing the measure, which has been bitterly opposed by leaders of the scientific community (see *Nature* 397, 459; 1999).

It was a similar amendment to last year's version of the bill by Senator Richard Shelby (Republican, Alabama) that implemented the Freedom of Information Act rule in the first place.

Opponents of the measure are backing the use of an amendment because they fear that a separate bill to overturn the measure, sponsored by George Brown (Democrat,

California) and Vernon Ehlers (Republican, Michigan), will not be given sufficient priority to pass into law.

Party marks anniversary of India's nuclear tests

[NEW DELHI] Scientists held nationwide celebrations for the first anniversary of India's nuclear tests last week. Prime minister Atal Behari Vajpayee, who had declared 11 May 'national technology day', told a meeting of scientists in New Delhi that the test carried out at Pokhran was a symbol of resurgent India. It "not only brought strength to national security but also self-confidence to our national mind," he said.

The prime minister, whose government lost a confidence vote in parliament last month, said that his government's strategy had been vindicated by events following the nuclear tests. "Economic sanctions failed to frighten us; technology denials failed to threaten us," he said, adding that Indian scientists deserved to be congratulated for this (see *Nature* 393, 197–198; 1998).

US plan for consortium to seek AIDS vaccine

[WASHINGTON] The new director of the US National Institutes of Health's Vaccine

Research Center is proposing that a consortium of pharmaceutical and biotechnology companies should be set up under the centre's auspices to hasten the development of an AIDS vaccine.

According to Gary Nabel, the consortium would not involve pooling of financial resources, but would provide a forum for cooperative research and for addressing obstacles to private sector vaccine development. "The idea is to break down barriers and to encourage collaboration [to] get everyone to the point where we have vaccine candidates that they can market."

Nabel says he envisions joint research in areas such as assessing vaccine vector effectiveness. Ultimately, the consortium might be enlisted to pursue a vaccine candidate where no single company had an interest in doing so. He marks current efforts by private companies to develop a vaccine as "two out of ten," and adds: "We want to get them well into the higher numbers".

Telemedicine think-tank set up by Allègre

[PARIS] France's ministry of higher education, research and technology has set up a scientific committee for a coordinated initiative on "telemedicine and health technologies". The committee's remit will cover databases,

medical imaging and computer-assisted diagnosis, as well as telemedicine itself. Claude Allègre, the science minister, is an enthusiast of this sector, arguing that it will profoundly transform medicine and create a lucrative industrial market.

The committee is chaired by Alain Carpentier, from the Hôpital Broussais in Paris, and is made up of physicians, physicists, mathematicians and computer scientists. The initiative is being led by Jacques Demongeot of the Université Joseph Fourier in Grenoble.

Science minister tipped to avoid Yeltsin's purge

[MOSCOW] Russia's science and technology minister, Mikhail Kirpichnikov, is being tipped to retain his post, despite being sacked from President Boris Yeltsin's cabinet last week. The new prime minister, Sergei Stepashin, says he does not plan major changes to the core of the cabinet of former prime minister Yevgeny Primakov.

Kirpichnikov is a familiar figure in the White House — the home of the Russian government — because he headed its science, culture and education department for several years.

He is also said to have benefited from the fact that he has not been particularly visible

during his relatively short ministerial career, and that science is currently considered a low priority by the government.

UK lecturers vote to strike for higher salaries

[LONDON] British university lecturers have voted for strike action in support of a 10 per cent pay claim, having rejected a 3.5 per cent pay increase offered by universities. The move was supported by 58 per cent of members of the Association of University Teachers who voted in a ballot, and could lead to a series of one-day strikes disrupting universities over the summer months.

The union says that since 1981 national average earnings have risen by 40 per cent in real terms, yet higher education salaries have risen by only 3.1 per cent. Union leaders say that a 3.5 per cent rise is not enough to begin to close the huge pay gap. Earlier this year university vice-chancellors received an increase of nearly 7 per cent. The first strike has been called for 25 May.

Luc Montagnier's Paris centre goes bankrupt

[PARIS] The Luc Montagnier Centre in Paris, named after the head of the French group that discovered the AIDS virus in the mid-

1980s, is to go into receivership just three years after it opened. The centre, created by Montagnier, was intended to serve as an outpatient clinic for HIV sero-positives as well as a research centre.

The centre was set up mainly with money from Sidaction, a television fundraising evening. But further funding — expected to come largely from donations — has not materialized. By going into receivership, Montagnier may draw attention to his plight and attract new funds. His centres in Rome, Abidjan, Ivory Coast, and New York are unaffected.

China's nuclear weapons now a tourist attraction

[BEIJING] A nuclear weapons research institute, the China Engineering Physics Institute, has been included on a science and technology tourism route that is being promoted by Mianyang City in China's Sichuan Province.

The institute is known to be one of China's main research centres for the development of strategic weapons. According to Xinhua news agency, the tourist route — the first in China — will also include factories that can be visited, including one owned by the Changhong Group, the fourth largest manufacturer of televisions in the world.



The full text of items on this page — and further details and background information about the Unesco/ICSU World Conference on Science, to be held in Budapest from 26 June to 1 July — can be found on *Nature's* website at <http://www.nature.com>

India plans to demand compensation for effect of brain drain

[NEW DELHI] India's delegation to the World Conference on Science plans to argue for compensation for the loss of trained manpower from developing to developed countries, according to Valangiman Ramamurthi, secretary of the Department of Science and Technology.

He says the delegation will also stress the need to protect unwritten traditional knowledge systems. It will seek funding from Unesco for critical areas of science of economic importance, as well as for regional networking of existing facilities.

The conference will be attended by an eight-member official delegation from Indian ministries of education and science, led by science minister Murli Manohar Joshi. "We expect that the non-official delegation — from science academies, universities and private organizations — will be much bigger," says Ramamurthi.

Despite the government's unhappiness with the NATO bombing of Serbia and the handling of the Kosovo crisis by the United States, officials in New Delhi say that this is unlikely to influence their participation in the conference. "We in the science ministry have been preparing seriously for this event and planning to play an active role," says Ramamurthi, adding that the Indian strategy for the meeting will be finalized next week.

He says India is pleased that some of the suggestions made at a preparatory meeting in Bangalore have been included in the latest revision of the proposed *Framework for Action* to be adopted at the conference.

Full text: <http://helix.nature.com/wcs/a34.html>

University women seek a central role at Budapest

[LONDON] A major international organization concerned with the position of women in universities has criticized the draft of a declaration due to be adopted at the World Conference on Science, arguing that it makes too little reference to issues relating to women in science.

The International Federation of University Women (IFUW), a non-profit organization of more than 180,000 women graduates, says women's issues are in danger of being marginalized at the conference by being allocated to a special afternoon workshop.

But Unesco officials reply that the *Framework for Action* document to be adopted at the meeting gives a detailed treatment of the issue. Anna-Maria Cetto says a dedicated session was chosen because the issue warranted "special attention" and argues that "you need to be proactive to change a situation". She adds: "It hasn't meant marginalizing women, you can be sure of that."

Muriell Joye, IFUW secretary general, says: "Our experience is that it is very dangerous to have a special event for women." The IFUW, which brings together 66 national federations and associations, was closely involved in Unesco's World Conference of Higher Education last summer.

"At the World Conference of Higher Education, only women attended [a similar] session, and the point was lost," says Joye. "When issues [affecting women], such as age and research grants, were raised in conference, we received comments such as 'you have your panel for that'. We were pushed aside."

Joye says women are barely mentioned in the draft declaration and in a document submitted to Unesco and the International Council of Science (ICSU) as part of the preparatory discussion for the Budapest meeting. IFUW says that "too often women are linked to, or even described as, minorities", although they make up more than half the population. Joye adds that the organization does not have the resources to provide a detailed response to the lengthy framework document.

"Marginalizing women at the conference by mention in a single paragraph, and a special afternoon workshop on science and gender, will only help perpetuate the misunderstanding that the lack of women in science is a curiosity but of little general importance," says the IFUW statement. "It must be seen as a failure to develop the potential of half the possible scientific resource base."

Cetto says having a special session has allowed preparatory meetings on the topic of women in science in the run-up to the conference, which have made "productive progress". "There is now much more clarity of what the issues are. I think the [women in science] session will be better for the preparatory work. It has enabled comparisons between Africa, Europe and Latin America."

"All the comments received until two days ago are being taken note of," says Cetto. She says she has not seen the list of responses yet but says that organizations such as the Third World Organization for Women in Science have been "close to the process".

Full text: <http://helix.nature.com/wcs/a32a.html>

Access to scientific knowledge 'should be a basic human right'

[MONTEVIDEO] Scientific and technological literacy is a "social and ethical right of all human beings", according to Eduardo Martinez, a regional specialist in science and technology management at the Unesco office in Montevideo, Uruguay.

He argues that one of the major challenges facing developing countries is therefore to make science and technology an essential part of the culture. "The popularization of science and technology must make such knowledge a central component of culture, of social awareness and of collective intelligence."

Martinez points out that the technical possibilities of gaining access to

information are changing our vision of the world. "Today, access to knowledge is synonymous with development, well-being and quality of life."

He argues that the areas to be reached by science and technology must be broadened to integrate formal education and communication with informal efforts in both fields, academic discourse with colloquial language, and laboratory materials with domestic objects and ordinary daily achievements.

Activities leading to the popularization of science and technology must be based on interdisciplinary dialogue and work, integrating diverse fields of knowledge and

different theoretical and methodological approaches, he says.

Martinez points out that in the past few years many interdisciplinary teams have been formed in Latin America to tackle the problem of communicating science and technology to large audiences. Many isolated experiments are now coordinated through the Network for the Popularization of Science and Technology in Latin America and the Caribbean (Red-POP).

Red-POP embraces more than 70 science and technology popularization centres and programmes. Its next biennial meeting will be held in June in Rio de Janeiro.

Full text: <http://helix.nature.com/wcs/c16.html>

Stand up for the rights of professors too

Sir — I have seen many letters in your pages, most recently from Eugen Tarnow (*Nature* 398, 657; 1999), suggesting solutions to the problem of supervisors and others unworthily attaching their names to the work of young scientists. But the opposite frequently occurs, so in seeking a solution the pendulum should not swing too far the other way.

In my field, senior scientists might be principal investigators for large space experiments requiring a decade or more of soul-destroying effort. At the end of that time there is usually a flood of new data which investigators and their students and postdocs analyse and publish, generally with the investigator's name attached in

recognition that their data are being presented for the first time, whether or not they participated in the analysis. I am sure that analogous situations exist in most research disciplines.

The problem arises when the newcomer publishes such data under their sole authorship, leaving bereft the principal investigator and others in the team, who may be under as much pressure as the youngster to show proof of research activity. Of course, a professor might be in a better position than a student to avoid abuse, but this is not always as easy as might be assumed.

In one extreme example, one of our students provided me with a slide for a

presentation that had been emblazoned with her personal copyright statement in large red letters. More commonly, I see papers submitted for publication without the names of those whose significant, unpublished output I know they have used.

The best safeguard against either problem is the integrity of those involved. But students and postdocs can sometimes be poor judges of what is appropriate, and should not be too strongly encouraged to feel they are being exploited when they may just be being nurtured.

F. W. Taylor

*Atmospheric, Oceanic and Planetary Physics,
Clarendon Laboratory, University of Oxford,
Parks Road, Oxford OX1 3PU, UK*

Grim reality of war... and plans for the peace

Sir — I was amused, in a grim sort of way, to read that scientists in Serbia have asked their colleagues in universities of NATO countries to oppose the bombardment of Serbia and Kosovo (*Nature* 398, 549; 1999). They point out that many Serb scientists opposed the Milosevic government's attack on the independence of their universities, and helped to force it to accept defeat in elections in 1996.

Well, good for them. But from this report it does not seem that they understand the reason for the NATO onslaught, which has nothing to do with the independence of their universities or the validity of their election process. The reason is the Milosevic government's pursuit of the evil policy of 'ethnic cleansing', the worst offence against humanity in Europe since Hitler's 'ethnic cleansing' of the Jews.

I am not a scientist, but I am opposed to the NATO bombing campaign. I hope that the scientists in Serbia are in turn opposed to the Milosevic government's actions against the Kosovo Albanians.

Ian Blake

Nibelungengasse 1, Vienna, Austria

Sir — Your report has reinforced my feeling that as scientists we need to show solidarity towards our colleagues in Serbia.

I think that a website should be set up where scientists could exchange views on the role they could play during this terrible crisis. And a fund of money and/or equipment should be donated to Serb scientists when the war is over. Presumably, basic research will not receive priority in the reconstruction of the country.

At times like this, even simple actions can relieve some of the pain. I would be happy to volunteer to assist in these tasks.

Stefano Casalotti

*Institute of Laryngology and Otology,
University College London,
330 Gray's Inn Road, London WC1X 8EE, UK*

Cancer fellowships awarded on merit

Sir — Stefano Parodi claims that he was treated unfairly by the fellowship selection committee of the International Agency for Research on Cancer (IARC) (*Nature* 398, 455; 1999). I served on the committee for many years, and I can report that it ranks its candidates as accurately as possible.

For example, it checks that there is no temporal trend in marking during the course of the two-day meeting, and that the final ranking is not significantly altered if the lowest and highest marks for each candidate are excluded (to avoid strategic marking). Lastly, the cut-off point for awarding fellowships is made where a significant gap in marks separates neighbouring candidates.

The real constraint is that there is only enough money to fund about 20% of candidates, even though, in my opinion, many show more promise than I did at their age. Parodi wants to see the list of candidates and their scores. That would be most improper. Would he really want his own name and score to be made public?

John Cairns

*Clinical Trial Service Unit, Radcliffe Infirmary,
Harkness Building, Oxford OX2 6HE, UK*

Sir — The IARC programme of research training fellowships began in 1967, since

when some 500 fellowships have been awarded (<http://www.iarc.fr>). The selection procedure aims to choose the most talented fellows, based on scientific merit.

After an interview of the candidate and a review of the application, priority scores are assigned by each member of the selection committee, which is an international group of 12 distinguished scientists with expertise in various cancer research disciplines. Some 10–12 of the 50–60 candidates evaluated each year are awarded fellowships, based on their average priority scores and the availability of funds. This process takes about four months, and all the candidates are informed soon after.

Parodi was interviewed by an epidemiologist speaking his native language, and an expert biostatistician reviewed his application. He was not among those selected, nor did he qualify as an alternative in the event of an award not being accepted. The selection committee met on 23–24 April 1998, and the candidate was notified of its decision by e-mail, fax and mail by the second week of May.

In June, he requested the list of all candidates, showing who had been accepted, the scores assigned, the reasons for exclusion and the criteria employed. We felt that it would be inappropriate to disclose this, and we are not aware of any fellowship programme that divulges such confidential data.

Ruggero Montesano

*Cancer Research Fellowship Programme,
International Agency for Research on Cancer, 150
Cours Albert-Thomas, 69372 Lyon Cedex 08, France*

Norman Breslow

*(Chairman, IARC Fellowships Selection
Committee)
Department of Biostatistics, UW.H656,
University of Washington, Box 357232,
Seattle, Washington 98195, USA*

What price ergonomics?

Ergonomists have a say in the design of almost everything in the modern world, but there is little evidence that their methods actually work. Here is an evaluation of those methods and of the worth of ergonomics in design.

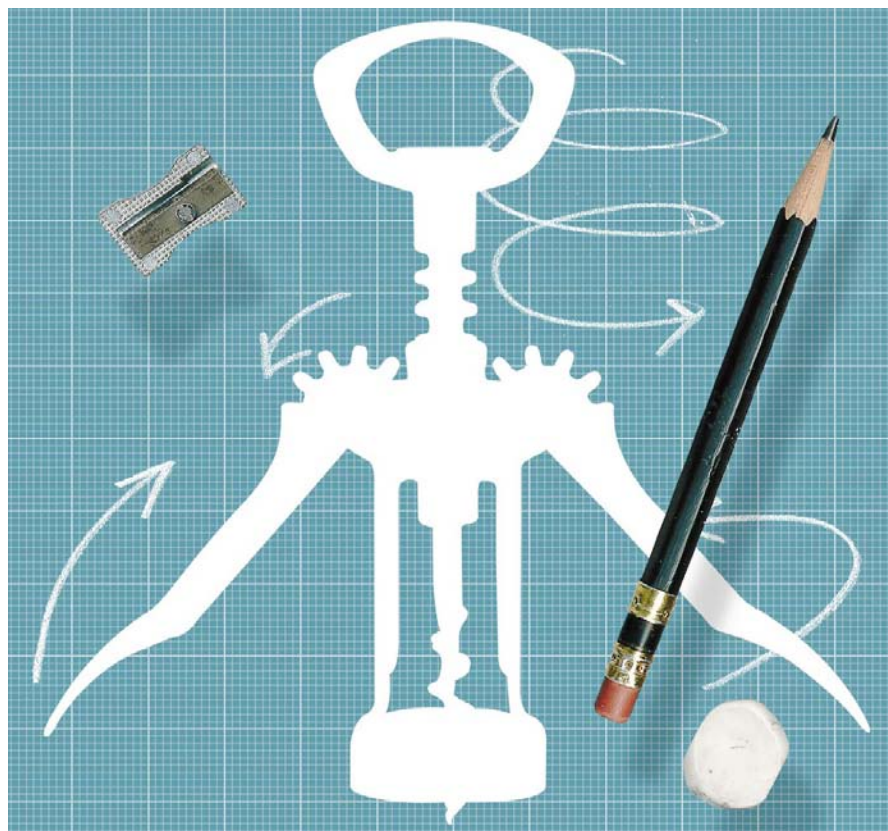
Neville A. Stanton and Mark S. Young

Engineers make things that are useful to people. In collaboration with designers, ergonomists make things that are usable by people. The concept of usability means making artefacts easy, efficient and comfortable to use (anything from a corkscrew¹ to a control room in a nuclear power station²). Most people have experience of poorly designed objects. At best they cause frustration and annoyance (for example when a video recorder fails to record your favourite programme). At worst they can lead to injury or even death (as in the release of radioactive material from a nuclear reactor).

'User friendly', 'easy to use' and 'ergonomically designed' are commonly found claims for domestic appliances and devices, particularly in the advertising business. For instance, an advertisement for a well-known music system boasts that it is 'Technologically advanced, yet remarkably easy to use'. But only a few people — including some, but not all, of those behind such advertisements — know what is involved in an ergonomic assessment.

Ergonomists do more than simply design chairs (a common misconception). Rather they are involved in various aspects of the design and evaluation of a host of products. They work on the specification of user requirements, the development of design guidelines, the evaluation of device prototypes, and the implementation and analysis of user trials. Traditionally, ergonomics has had most influence on the physical characteristics of design. For example, in conceiving of a new radio-cassette machine, ergonomists might advise on the buttons: their size and placing, and the reach and force required to activate them. The ergonomist draws on a substantial body of research findings and hard data, such as human anthropometrics — tables of data that contain information about the range of human physical attributes and characteristics³.

Nowadays, ergonomists also advise on interactive devices, such as so-called WIMP (Windows, Icons, Menus and Pointers) interfaces. Advice includes the form that the interface should take, relating the grouping of functions and device elements to the goals of the user, predicting user satisfaction, and estimating task performance time and the error potential of the device. These analyses are used to select from a choice of alternative designs and to improve the chosen option. But analyses of human cognitive functioning do not rest upon hard data to the same ex-



tent as anthropometrics⁴. Questioning these techniques is our purpose here.

Despite the proliferation of ergonomic methods in research⁵, teaching⁶ and industry⁷, there is little substantive evidence that these methods actually work⁸. Part of the reason is that researchers tend to stick to one or two that they know, trust and, in many cases, have developed themselves. The main point here, however, is that methods for predicting the usability of devices have become so entrenched that their validity is simply assumed and seldom tested. These methods are used in good faith, in the belief that the results are reliable and valid. But that assumption is questionable, and we find that some approaches do not do all that is claimed for them.

Reliability and validity

The objective way to see whether ergonomic methods work is to assess their reliability and validity. Methods can be reliable without being valid, but they cannot be valid without being reliable. We can liken reliability and validity to the accuracy of a rifle marksman using a gunsight. Reliability refers to the grouping of the shots (the error of the marksman), whereas validity refers to the closeness of each shot to the centre of the target (the error of the gunsight). If ergonomic methods

are shown to be adequately reliable and valid, they may be used with increased confidence.

In a series of studies, we took ten of the more popular ergonomic methods, and subjected them to tests of reliability and validity (details of the methods are found in ref. 1, and ergonomic textbooks). The ergonomic methods each predicted different aspects of usability, from objective performance measures (such as reaction time and errors) through to subjective criteria (such as user satisfaction). The methods evaluated included classical approaches (such as interviews and observation), as well as some of the more modern techniques (such as SHERPA, the Systematic Human Error Reduction and Prediction Approach). Only a few people were involved in the study: eight were trained in each of the methods and 30 were observed using a radio-cassette machine, the device we chose to use. But these numbers are realistic in terms of the kind of studies carried out by ergonomists. Our analyses produced between 500 and 1,650 data points for each method, depending upon the method used, which were classified as hits, misses, false alarms and correct rejections.

The results, which appear in Table 1, show huge variability in both reliability and validity of the various methods. The validity

statistics were calculated using either Pearson or Phi (a correlation coefficient for dichotomous data). For the latter PhiMax is also given, as this shows the maximum value that could be achieved, given the marginal distributions⁹. Interestingly, two of the methods return small but negative correlations, meaning that the predicted behaviour was the opposite of that observed.

So is it worth using these methods at all? Thus far in the argument, the answer has to remain 'perhaps'. The reason is that ergonomists not only have an agenda of making life in our technological world easier, safer and more enjoyable, but also that of helping to maximize manufacturer revenue and minimize costs. Designing a product in accordance with usability principles can increase profits; and, given the sums of money involved in product development, even methods with relatively low validity could offer returns.

Methods that can predict good or bad designs should save time and money in product development. They might increase sales and profits, and reduce need for training and after-sales support. But those benefits have to be set against the costs of setting up an ergonomics laboratory, conducting the methods and evaluating iterative prototypes. So does the relative inaccuracy of the methods reduce these benefits beyond the point where they are worth the effort and cost? That is, does it cost more to use a method than is gained from the end result?

Cost-benefit analysis

Claims as to the cost-effectiveness of ergonomic methods normally centre on the results of cost-benefit analysis. Here the approach is to calculate the cost of applying the method (in terms of person-hours, materials and so on) and subtract this from the estimated savings generated by the improvement in design. The net figure is proposed as the benefit brought about by using ergonomic methods.

The figures can look good. According to one textbook example¹⁰, the cost of a usability assessment of the interface for new computer software (US\$171,445) compares very well to the potential savings of the new

design (\$592,635) in terms of the reduced costs of training and after-sales support in the first year. On the face of it, then, in this instance ergonomic methods provide a profit of more than \$400,000.

The estimate of likely economic benefits does not however attempt to take the accuracy of the ergonomic methods into account. The application of ergonomic methods (if inaccurate) could fail to predict true problems with device design or, worse still, identify problems that do not exist. We argue that the degree to which the financial benefits will be realized will depend upon the accuracy of the methods used. Using the reliability and validity data we acquired, the net benefit of each method can be calculated from its utility U (that is, its practical usefulness couched in financial terms): $U = (\text{validity} \times \text{prospective increase in value}) - \text{costs}$. A recalculation of the textbook cost-benefit analysis would find a substantial reduction in the benefit by including the validity data. It is therefore essential that these data are considered when addressing utility, as they show how much variability can be accounted for in ergonomic methods.

The practical context here is that, in product design, a manufacturer is typically faced with the prospect of choosing between ten or more designs. Given this degree of choice, how are they to arrive at the best decision? They could flip a coin or rely upon intuition. Alternatively, they could use the methods to predict differences in user performance to select the best design. The difference between a good and poor design could be worth millions of dollars to a manufacturer over the lifetime of a product.

For the purpose of testing the formula with the computer software example¹⁰, we can estimate that choosing a good design over an average design will increase profitability by upwards of \$500,000 per year (a rounding down of the over-precise figure provided by the example). Similarly, we use \$100,000 as the cost of making prototypes and assessing the device. The same costs are used in both analyses to show the effect of validity on utility. In our analysis, the method with the lowest validity value was use of repertory grids; that with the highest

was the keystroke level model (KLM). The contrast of utilities brought about by these two methods, with respective validity values of 0.078 and 0.769, can be estimated as: $U(\text{repertory grids}) = (0.078 \times \$500,000) - \$100,000 = -\$61,000$; $U(\text{KLM}) = (0.769 \times \$500,000) - \$100,000 = \$284,500$. So some methods may return negative values in the first year, whereas others show a healthy return on investment.

A range of factors can affect the selection of ergonomic methods to be applied in evaluating specific products, including time and resources available, efficacy of the methods, criteria by which the product is to be evaluated, and access to a 'test' sample of end-users. Utility analysis can help make some of these judgements more explicit and more objective. It can help in the selection of methods, show where effort may be best expended, and make the design process more effective. Experience from the field of personnel selection shows that for utility analysis to be effective it must be easy to use and the derivation must be transparent — people using the formula must be able to see how it works and carry out the calculations without having a PhD in mathematics.

We can provide the reliability and validity data, the costs of conducting each method, and the appropriate procedure for conducting a cost-benefit analysis. Product developers will have their own budgets for sales and training, and these will differ depending on whether the product is a vending machine or a video recorder. So it is up to the product development teams concerned to make their own calculations and decide whether ergonomic methods are actually worth using.

Our aim here has not been to condemn ergonomic methods or ergonomics as a discipline. Rather, our take-home message is that, if ergonomists are really to improve the design of artefacts, their methods have to be shown to be both useful and usable. □

Neville A. Stanton and Mark S. Young are in the Engineering Psychology Research Group, Department of Psychology, University of Southampton, Southampton SO17 1BJ, UK.
e-mail: n.a.stanton@soton.ac.uk

Table 1 Reliability and validity of ergonomic methods

Ergonomic method	Reliability	Validity	PhiMax
Checklists	0.307	0.206*	0.659
Heuristics	0.471	0.087*	0.464
Hierarchical task analysis	0.226	0.188*	0.937
Interviews	0.449	-0.112*	0.422
Keystroke level model	0.916	0.769*	N/A
Link analysis	0.830	0.356*	0.572
Observation - errors	0.890	-0.141*	0.343
Observation - time	0.890	0.729*	N/A
Questionnaires	0.578	0.615*	N/A
Repertory grids	0.562	0.078	0.681
SHERPA	0.392	0.238*	0.922

Reliability was calculated by Pearson's product moment correlation coefficient. Validity was calculated by Phi (those with associated PhiMax values) and Pearson's product moment correlation coefficient. Asterisk indicates that validity value was statistically significant.

1. Stanton, N. A. *Human Factors in Consumer Products* (Taylor & Francis, London, 1998).
2. Stanton, N. A. *Human Factors in Nuclear Safety* (Taylor & Francis, London, 1996).
3. Pheasant, S. T. *Bodyspace — Anthropometry, Ergonomics and Design* (Taylor & Francis, London, 1986).
4. Dowell, J. & Long, J. T. *Ergonomics* **41**, 126-139 (1998).
5. Salvendy, G. *Handbook of Human Factors and Ergonomics*, 2nd edn (Wiley, New York, 1997).
6. Wilson, J. & Corlett, N. *Evaluation of Human Work*, 2nd edn (Taylor & Francis, London, 1995).
7. Jordan, P. W., Thomas, B., Weerdmeester, B. A. & McClelland, I. L. *Usability Evaluation in Industry* (Taylor & Francis, London, 1996).
8. Stanton, N. A. & Young, M. *Appl. Ergonomics* **29**, 41-54 (1998).
9. Guilford, J. P. *Psychometric Methods* (McGraw-Hill, New York, 1957).
10. Bias, R. G. & Meyhew, D. J. *Cost-Justifying Usability* (Academic, Boston, 1994).

Acknowledgements. The research reported here was supported by the EPSRC under the LINK transport initiative. We thank S. F. Blinkhorn for advice on the content and form of this article.

A shared but complex bridge

Robert E. Kingston

The first step in the 'expression' of genes to form proteins is transcription, the process by which DNA is converted to an RNA template for the protein-producing machinery. A hitherto unknown player in transcription of human genes is now emerging.

The human genome has roughly 100,000 genes, the expression of which must be exquisitely regulated. Much of this regulation is directed by activators — proteins that bind to specific DNA sequences and activate gene transcription by an enzyme called RNA polymerase. In mammalian cells there are hundreds of activators, and a small subset of these binds to each gene and responds to developmental or environmental cues. These activators have been categorized, and the proteins that transcribe DNA to RNA dissected, so the transcription field should now be more or less wrapped up (Fig. 1). But it's not as simple as that. Activators often do not interact with the RNA polymerase directly; instead, they use protein complexes known as 'coactivators' or 'mediators', which relay biological signals from the activator to the polymerase to turn on transcription. Several papers^{1–7} (the latest of which, by Boyer *et al.*¹, appears on page 276

of this issue) report the identification of such mediator complexes in mammalian cells, and reveal some surprising characteristics.

Mediator complexes are much larger than common sense would dictate. In mammals they vary from nine to 20 or more peptides, with a relative molecular mass ranging from 600,000 to over 1,000,000. Yeast media-

tor complexes are also quite large^{8,9}, and contain many proteins that have direct equivalents in mammals. Why are so many proteins needed to relay a signal from an activator to RNA polymerase? It would be ungainly to use a different complex for each of the many different activators, so how many different mediators are there? The recent papers begin to address these questions.

The first finding is that different activators can interact with the same mediator complex. Rachez *et al.*², Näär *et al.*³ and Ito *et al.*⁴ show that four mediator complexes presumed to be different — I'll spare readers their full names here, and call them TRAP, DRIP, ARC and SMCC — are, in fact, identical. The DRIP and TRAP mediators were discovered through their interactions with activators that respond to hormones, so it is no surprise that they are identical. It also makes sense that SMCC and ARC are one and the same, because they can both act with the

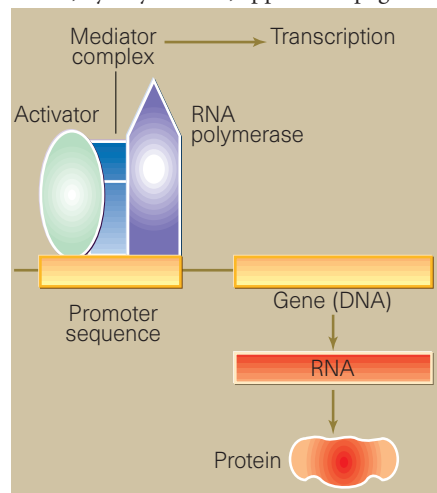


Figure 1 Some of the basics of transcription. An activator protein binds to a specific region of DNA, known as the promoter, just ahead of the gene that is to be transcribed. This activator turns on an enzyme, RNA polymerase, which tracks along the gene, copying the DNA template into RNA. The activator does not always contact the polymerase directly, and the activating signal may be relayed through a mediator complex, which is made up of many different protein subunits. The RNA product of transcription itself serves as a template for the protein-producing machinery.

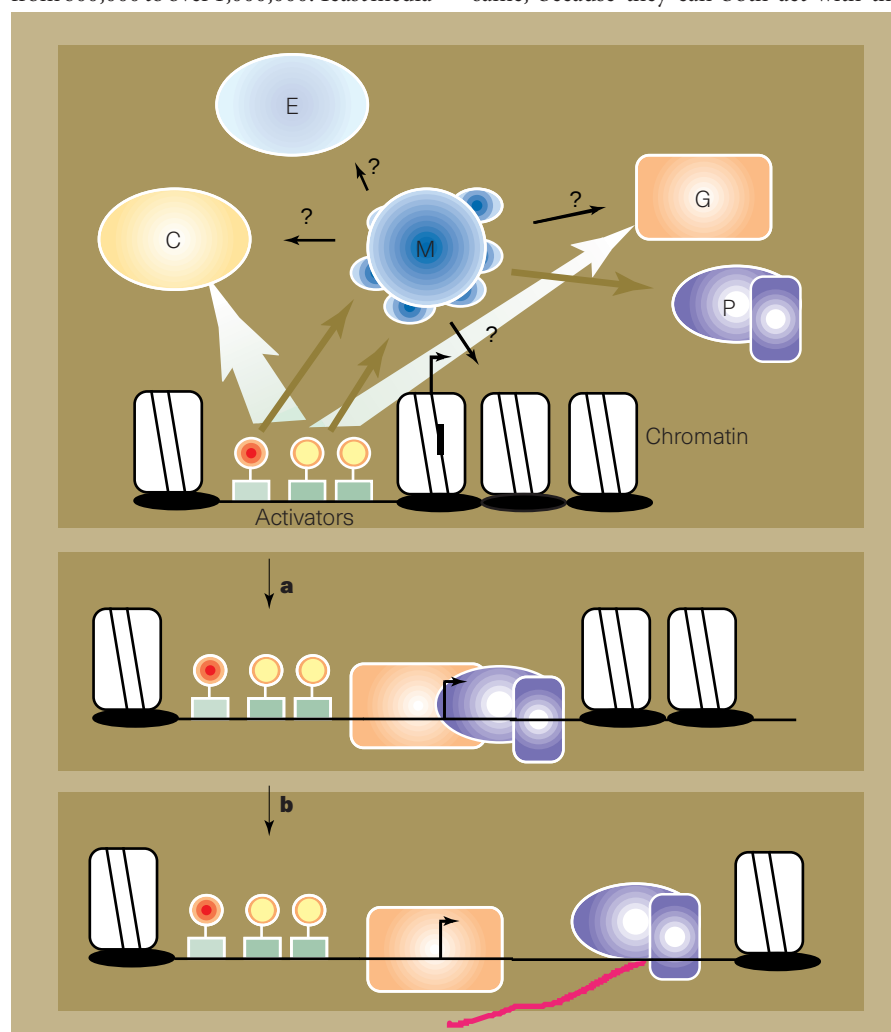


Figure 2 Transcription in more detail, and potential targets for mediator complexes. The activators touch many sites on mediators (M), which contact RNA polymerase (P) and might also contact the machinery that unravels chromatin (C), the general transcription factors (G), elongation factors (E) or chromatin itself. Activators also directly contact the general transcription machinery and chromatin-modifying proteins. The protein-protein contacts might modulate any of the steps in transcriptional activation. These include: a, unravelling of the chromatin and recruitment of general transcription factors and RNA polymerase; and b, opening of the DNA helix, followed by clearance of the promoter by RNA polymerase and elongation through the gene.

same activator. But the apparent identity of DRIP/TRAP with ARC/SMCC was surprising, as the activators that require these mediators have diverse biological functions.

A second characteristic of mammalian mediators is that many of the individual proteins (or subunits) within each mediator complex are shared between the different complexes. For example, the yeast MED7 protein⁹ is a subunit that belongs to at least five different mediator complexes (human Mediator¹, mouse Mediator⁶, NAT⁷, CRSP⁵ and TRAP/DRIP/ARC/SMCC²⁻⁴). Several other subunits are also shared. These common subunits aside, complexes otherwise have different compositions and abilities to interact with various activators. Many of the subunits that are found in several different complexes can interact with RNA polymerase, supporting the idea that, in mammals, mediators provide a direct link between activators and RNA polymerase.

This sharing of protein subunits between different mediator complexes provides a tidy answer to how so many large complexes can be encoded efficiently. If, for example, the genome encoded 40 proteins used in mediators, these could be divided into only four different ten-subunit complexes if no proteins were shared. But by mixing and matching subunits, a huge number of different complexes can be created. Moreover, the concept of shared subunits has been seen in other proteins involved in transcriptional regulation. For instance, the so-called TBP-associated factors are shared between one group of complexes that bind to a specific DNA sequence, and another group which interact with chromatin (the form in which the DNA is packaged in chromosomes)^{10,11}.

Why are mediator complexes so large? Their size almost certainly reflects the need to interact with diverse activators. For example, Boyer *et al.*¹ and Sun *et al.*⁷ report that the human equivalent of Sur-2 — a protein originally identified in the nematode worm *Caenorhabditis elegans* — belongs to both the human Mediator and NAT complexes. Boyer *et al.* further show that Sur-2 seems to interact with some, but not all, of the activators that use human Mediator. The activators that function through human Sur-2 are involved in similar signalling pathways. Moreover, when the *Sur-2* gene is mutated in worms, several different developmental processes are affected. So, perhaps Sur-2 is an essential link in the mechanisms that increase transcription in response to developmental signals.

The size of the mediator complexes probably also reflects the need to regulate many different steps in transcription (Fig. 2). Most activators control more than one step, and they can also interact with other complexes involved in transcription. Although the mediators probably do not regulate all of these steps, they may, nonetheless, make

numerous regulatory contacts, meaning that they need to be large.

Another layer of complexity is that DNA is not naked in the cell — it is wrapped around another group of proteins to form chromatin, which must be unravelled before the genes can be transcribed. By comparing DRIP/ARC with other mediators, Rachez *et al.*² and Näär *et al.*³ illustrate the different functions of different mediators with regard to this chromatin structure. They see no stimulation of DRIP/ARC-mediated transcription on naked DNA, but find that transcription is stimulated on chromatin using several different activators. The authors conclude that DRIP/ARC regulates a step that is particularly important when the DNA is wound up as chromatin. The explanation could be that DRIP/ARC directs unravelling of the chromatin. Or, it might increase the rate of a step (such as elongation of the RNA product by RNA polymerase) which occurs without help on naked templates, but which requires assistance to work on chromatin.

This aspect of DRIP/ARC function distinguishes it from other mediators — such as CRSP and human Mediator — which are very active on naked DNA templates^{1,5}. However, Ito *et al.*⁴ find that TRAP/SMCC (which is identical to DRIP/ARC) can stimulate transcription on naked DNA templates with a variety of activators. This discrepancy probably reflects variations in the *in vitro* transcription systems used, which might allow stimulation on naked templates to be detected under one set of conditions, but not under another. The complex might, in fact, regulate several steps in transcription, the relative importance of which varies depending on the experimental system used.

The picture of mammalian mediators

that emerges from these studies is one of large complexes containing distinct sets of proteins that interact with activators, RNA polymerase or other factors, to regulate different steps in transcription. Only one protein in the mediator complex may bind to each activator, then another few proteins might effect a function while the rest of the proteins in the complex sit by and watch. With other activators, different proteins could be pressed into action. So, for a given gene, transcriptional activation might first require a set of activators to bind (Fig. 2). Next comes unravelling of the chromatin structure by large, chromatin-specific complexes, then recruitment of large mediator complexes and binding of other, general transcription factors. The final step is transcription by RNA polymerase. In this way there is an intricate, regulated dance of huge complexes on a small region of DNA — beautiful to behold but terrifyingly complicated. □

Robert E. Kingston is in the Department of Molecular Biology, Wellman 10, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

e-mail: kingston@frodo.mgh.harvard.edu

1. Boyer, T. G., Martin, M. E. D., Lees, E., Ricciardi, R. P. & Berk, A. J. *Nature* **399**, 276–279 (1999).
2. Rachez, C. *et al.* *Nature* **398**, 824–828 (1999).
3. Näär, A. M. *et al.* *Nature* **398**, 828–832 (1999).
4. Ito, M. *et al.* *Mol. Cell* **3**, 361–370 (1999).
5. Ryu, S., Zhou, S., Ladurner, A. G. & Tjian, R. *Nature* **397**, 446–450 (1999).
6. Jiang, Y.-W. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 8538–8543 (1998).
7. Sun, X. *et al.* *Mol. Cell* **2**, 213–222 (1998).
8. Myer, V. E. & Young, R. A. *J. Biol. Chem.* **273**, 27757–27760 (1998).
9. Kim, Y.-J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. *Cell* **77**, 599–608 (1994).
10. Ogrzyzko, V. V. *et al.* *Cell* **94**, 35–44 (1998).
11. Grant, P. A. *et al.* *Cell* **94**, 45–53 (1998).

Particle physics

Only a matter of time

Ken Peach

A steady stream of new results¹⁻⁴ over the past six months has illuminated an obscure but important corner of the quantum world, which could have profound implications for our understanding of the origin of the material universe. These experiments search for small asymmetries in particle properties associated with the behaviour of quantum-mechanical systems under the reversal of time, or, equivalently, the combined operation of charge conjugation (interchange of matter and antimatter) and parity (reflection of space through the origin).

Most of the symmetries of the quantum universe have their counterpart in the macroscopic universe. For example, the fact that physics is the same in London and New York leads directly to the conservation of momentum in both the quantum and classi-

cal worlds. Quantum effects become important only when the action (momentum applied for a given distance or energy applied for a given time) is comparable to Planck's constant. Because this is such a small quantity, quantum effects are negligible in most everyday circumstances. Quantum asymmetries become observable only when two or more different paths between the initial and final states can interfere.

The three fundamental quantum symmetries of charge conjugation (C), parity (P) and time reversal (T) have no classical analogue. Charge conjugation is clearly a quantum-mechanical concept — antimatter does not exist in the classical world. The need for, and existence of, antimatter is an inevitable consequence of the combination of (special) relativity and quantum mechanics. The

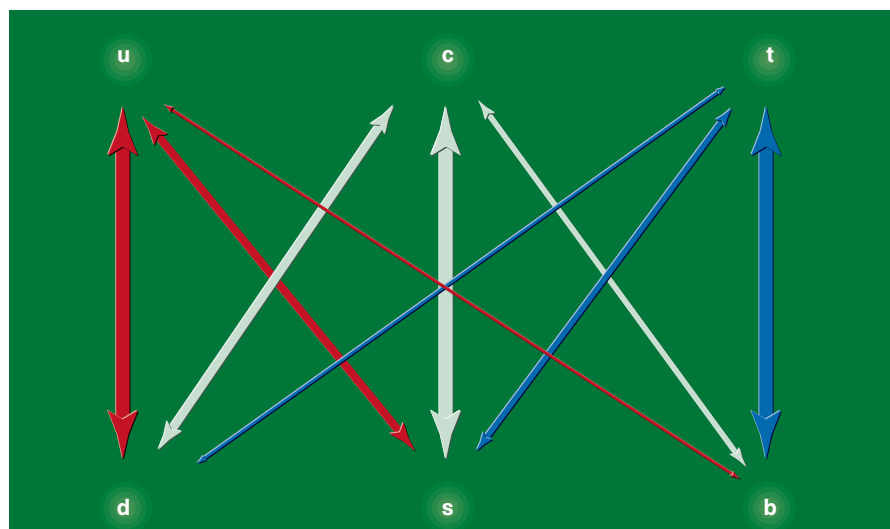


Figure 1 The weak interaction, responsible for nuclear β -decay, mixes the different quarks together, but there is a hierarchy in the strength of the interactions. For example, the up quark interacts mostly with the down quark, a little with the strange quark and hardly at all with the bottom quark (indicated by line thickness). The CP or T violation arises because the amplitudes of these transitions cannot all become simultaneously real if there are three or more generations of quark.

observation that matter dominates antimatter in the visible Universe is a clear indication that matter and antimatter are not completely interchangeable.

Parity is a more familiar concept — an ordinary mirror reverses one of the dimensions of space leaving the other two unchanged, whereas the ‘parity mirror’ reverses all three directions of space. As Lewis Carroll realized, the mirror world may look similar to the real world, but there is scope for subtle differences in the hidden domain. In the quantum world, this subtle difference arises because the underlying physics is described by a quantum-mechanical ‘amplitude’, whereas the observation in the real world corresponds to the ‘intensity’, or square of the amplitude. So, it is possible that the same observation (intensity distribution) arises from quantum-mechanical amplitudes that are subtly different. Observable asymmetries, such as parity violation in nuclear β -decay, occur when two of these subtly different amplitudes can be made to interfere with one another, and so produce an intensity pattern that depends on the degree of difference between them. The classic demonstration of parity violation in nuclear β -decay is made possible by the interference of parity-preserving and parity-changing amplitudes.

Time reversal also seems easy to understand — reversing the direction of time should be just like running a film backwards. At the level of particle interactions, this cannot lead to asymmetries, because simply running the film backwards must lead to the same initial state — this is known as the principle of reversibility. However, the time-reversal operation also means reversing the sign of the imaginary component of the quantum-mechanical amplitude. If there is a real and an imaginary component to the

amplitude, the application of the time-reversal operator will lead to an amplitude that is physically different.

Although these three operators seem very different in character, they are related in a fundamental way through the theorem of Pauli, Luders and Bell. They showed in the early 1950s that relativistic quantum-field theories (of which the highly successful ‘Standard Model’ of particle interactions is an example) must be invariant under the combined operations of C, P and T taken in any order. At the time, it was assumed that C, P and T were independently conserved by all of the fundamental forces. Indeed, they are conserved in the strong and electromagnetic interactions. However, the weak interaction — which governs radioactive β -decay and makes the Sun shine — has now been shown to violate all of the symmetries individually and in pair-wise combination. It is still assumed that the combination of all three is conserved by all three interactions, although it is possible that it could be violated by the gravitational force at the Planck scale.

Two experiments^{1,2}, reported late last year and discussed in News and Views⁵, observed the first direct evidence that nature — or at least particles called kaons — does not remain invariant under time reversal. Both experiments used the decay of neutral kaons to search for time non-invariance. Perhaps unsurprisingly, the level of violation observed was consistent with that expected from previous experiments showing CP violation in the neutral-kaon system.

Earlier this year, a new upper limit was reported³ on the electric dipole moment of the neutron. The existence of a permanent electric dipole moment along the spin axis of a particle is direct evidence for violation of time-reversal symmetry. The electric dipole

moment of the neutron is expected to be very much less than $10^{-32} e \text{ cm}$ (the units are such that the centre of negative charge is different from the centre of the positive charge by less than 10^{-32} cm when the charge is scaled to that of the electron, e). This expectation arises from the Standard Model of particle interactions believed to describe the CP violation and T violation observed in the neutral-kaon system. The observation that the electric dipole moment of the neutron is less than $6.3 \times 10^{-26} e \text{ cm}$ severely constrains various extensions to the Standard Model because these almost inevitably lead to a much larger electric dipole moment for the neutron.

Within the Standard Model, CP violation and T violation can be accommodated provided that there are at least three generations, or families, of quarks (see Fig. 1). Unlike the strong and electromagnetic interactions, which effectively only connect a particle to itself or its own anti-particle, the weak interaction mixes the generations, and connects up-like quarks (up, charm, top) to down-like quarks (down, strange, bottom). This process is called mixing. If the Standard Model does lead directly to CP violation through quark mixing, a small but measurable difference is expected in the amount of CP violation in the two different two-pion decay products of the kaon. This difference would help distinguish the Standard Model theory of CP violation from alternatives, such as the superweak interaction, which predicts that there should be no such difference.

The KTeV experiment at Fermilab has shown that the amount of CP violation in the two decay products is different. They reported⁴ that there should be $1.68 (\pm 0.25)\%$ less CP violation in the decays to two neutral pions compared with the decays to two charged pions, as might be expected in the Standard Model. The importance of this new result is that it provides a precise measurement of this small effect (CP-violating decays account for only 0.2% of all decays of the neutral kaon). Previous experiments disagreed about whether this effect had been observed. The NA31 collaboration at CERN (on which I worked) reported the effect more than ten years ago⁶, and the final result⁷ of $1.38 (\pm 0.39)\%$, published in 1993, is remarkably close to the new Fermilab result. But a contemporary result⁸ from the E731 experiment, also at Fermilab, reported an effect of $0.45 (\pm 0.35)\%$, consistent with no effect and in poor agreement with the CERN result. The latest Fermilab result confirms the earlier CERN observation. A result is expected soon from a new CERN experiment (NA48) which, if it is consistent with the NA31 and KTeV results, will establish the effect beyond doubt.

Although an effect as large as that reported by KTeV can be accommodated by the current calculations within the Standard Model, it requires that several parameters within the

calculation be at the edge of the range allowed by other observations. Taking more 'reasonable' values for these parameters predicts a much lower value for this effect. This could be a hint that we are on the threshold of physics just beyond the Standard Model. All of which makes for the prospect of truly spectacular results from the B-factories — about to start taking data at the Stanford Linear Accelerator, California, and at the KEK accelerator in Tsukuba, Japan. Large asymmetries within the Standard Model are expected for some decay modes of the B-meson. If the hint of new physics from these first results turns into

reality, we could soon be in for some surprises. Ken Peach is in the Department of Particle Physics, Rutherford Appleton Laboratory, Chilton, Oxfordshire OX11 0QX, UK.
e-mail: Ken.Peach@rl.ac.uk

1. Angelopoulos, A. *et al.* *Phys. Lett. B* **444**, 43–51 (1998).
2. Arenton, M. *Workshop on Heavy Quarks at Fixed Targets*, Fermi National Accelerator Laboratory, Chicago, 10–12 October 1998.
3. Harris, P. G. *et al.* *Phys. Rev. Lett.* **82**, 904–907 (1999).
4. Shawhan, P. *Fermilab Seminar*, Fermi National Accelerator Laboratory, Chicago, 24 February 1999; see <http://fnphx-www.fnal.gov/experiments/ktev/epsprime/epsprime.html>
5. Peach, K. *Nature* **396**, 407–409 (1998).
6. Burkhardt, H. *et al.* *Phys. Lett. B* **206**, 169 (1988).
7. Barr, G. D. *et al.* *Phys. Lett. B* **317**, 233–242 (1993).
8. Gibbons, L. K. *et al.* *Phys. Rev. Lett.* **70**, 1203–1206 (1993).

Cancer

Many vessels, faulty gene

William G. Kaelin Jr

Hereditary cancer syndromes can teach us much about the molecular circuits that are central to the life of a cancer cell. The genes responsible for these syndromes often encode proteins that act at critical points in the pathways controlling fundamental processes such as cell division, differentiation and death. So, any alterations in these genes may give rise to the changes associated with cancer. But cancerous cells must also interact with the surrounding tissue, securing an environment that allows a tumour to grow. Formation of new blood vessels (angiogenesis) seems to be a big

hurdle in this regard and yet, until recently, no hereditary cancer gene had been directly linked to this process. But the paper by Maxwell *et al.*¹ on page 271 of this issue, together with two reports in *Science*^{2,3}, provides a considerable insight into how one gene, the von Hippel–Lindau (VHL) gene, regulates angiogenesis.

The VHL hereditary cancer syndrome was first reported about 100 years ago, although cases dating back to the 1700s have been identified. Affected people develop blood-vessel tumours, called haemangioblastomas, which affect the retina, cerebel-

lum and spine. Patients may also develop tumours in their adrenal glands, as well as clear-cell kidney cancers. The VHL gene was isolated in 1993, and it conforms to Knudson's two-hit model. That is, people with the disease are born with one mutated copy of the VHL gene. Tumours develop when the other, normal, copy is inactivated — by mutation or hypermethylation — in a susceptible cell. People with two normal copies of the VHL gene may also develop sporadic haemangioblastomas and clear-cell kidney cancers if both copies are inactivated. In this case, the VHL gene is inactivated some time after birth.

Why does loss of the VHL gene give rise to tumours? Its protein product, pVHL, does not closely resemble any other protein. But it does form complexes⁴ with other proteins called elongin B, elongin C and Cul2 (Fig. 1). Elongin C and Cul2 are similar to two yeast proteins, Skp1 and Cdc53, respectively, which themselves form multi-protein complexes (known as SCFs, for Skp1/Cdc53/F-box protein). These complexes target other proteins in the cell for degradation via a process known as ubiquitination. The target is bound by a protein called ubiquitin, which labels it for destruction in the cavity of a protein-digesting complex called the proteasome. Because the F-box protein determines the substrate specificity of the SCF complex, pVHL might, by analogy, regulate protein turnover.

Stebbins *et al.*² have now solved the crystal structure of pVHL bound to elongins B and C, and found that there are indeed similarities to yeast SCF complexes. The pVHL contains two subdomains, referred to as alpha and beta. The alpha domain binds to elongin C which, in turn, binds to elongin B (ref. 2) and Cul2 (ref. 5). The beta domain is thought to be where the target protein binds. Both of these subdomains are often mutated in people with VHL disease. Elongin C, as predicted, resembles yeast Skp1; and the part of pVHL that binds elongin C is similar to an F box. Further evidence that pVHL is involved in ubiquitination comes from the finding, by Kamura *et al.*³, that pVHL copurifies with a protein called Rbx1 or ROC1. This protein helps the SCF to degrade substrate proteins by recruiting a ubiquitin-conjugating enzyme to the complex^{3,6–8}. Elongin B itself also contains a region that resembles ubiquitin, although there is no evidence that elongin B becomes covalently linked to proteins.

If pVHL regulates ubiquitination — or, perhaps, an analogous process — what are its targets? The tumours found in VHL patients contain many blood vessels, a property that has been linked to production of tumour-derived vascular endothelial growth factor (VEGF; also known as vascular permeability factor). Patients may also occasionally produce excess red blood cells⁴, a property

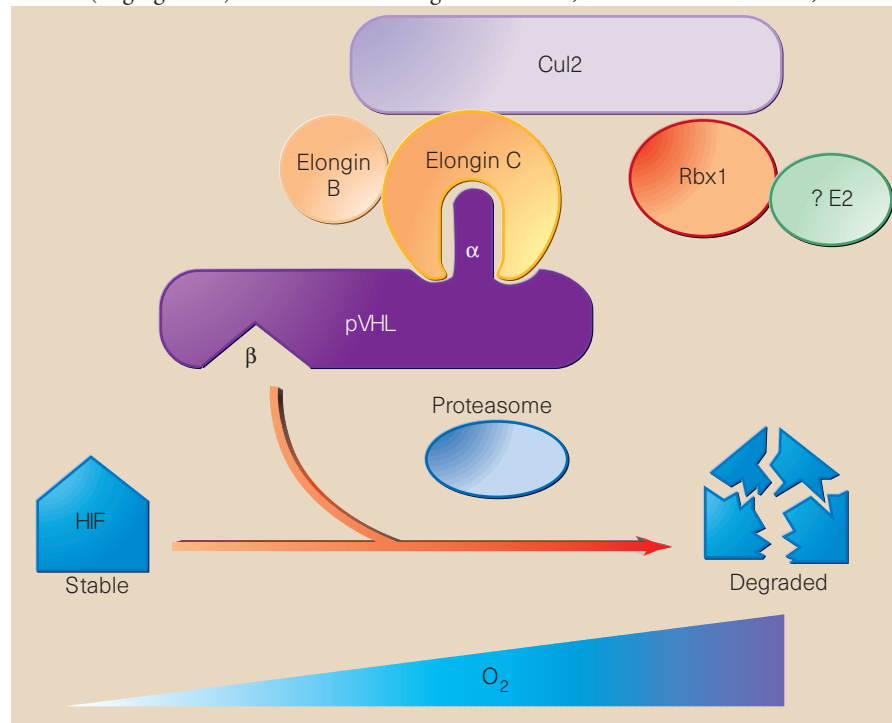


Figure 1 Molecular basis of von Hippel–Lindau disease. The VHL protein (pVHL), bound to elongin C, elongin B, Cul2 and Rbx1, degrades alpha subunits of hypoxia-inducible factor (HIF) in an oxygen-dependent manner. Maxwell and colleagues¹ have shown that cells lacking pVHL do not degrade HIF in response to oxygen. Because HIF normally controls the production of factors that promote formation of blood vessels, when it is not degraded the result is uncontrolled proliferation of blood vessels. E2, unknown ubiquitin-conjugating enzyme.



100 YEARS AGO

A report by Prof. David Hanseemann, of Berlin, on the brain of Hermann von Helmholtz, is referred to in the *British Medical Journal*. The great physicist died of apoplexy on September 8, 1894, at the age of seventy-three. The circumference of the head was 59 centimetres, that of the skull 55 centimetres. The breadth of the skull was 15.5, and its length 18.3 centimetres. The cephalic index was therefore 85.25, showing a broad head. Helmholtz's head was about equal in size to that of Bismarck, and rather smaller than that of Wagner, both of whom had big heads. On the other hand, Darwin's head was only 56.3 centimetres in circumference. The weight of the brain, with its blood, was 1700 grams, without the blood 1440 grams, being about 100 grams heavier than the average human brain.

From *Nature* 18 May 1899.

50 YEARS AGO

The Centenary of Bedford College for Women. Queen Victoria had been only twelve years on the throne when there was inaugurated in Bedford Square, London, a "Ladies' College" at which young women might acquire something more than their customary, somewhat vacuous, social background. Its founder, Mrs. Reid, had dreamed for many years of such a college for women, and it was owing to her ceaseless effort and zeal that Bedford College opened its doors a full twenty years before Girton College was opened at Hitchin or the idea of Newnham College was conceived. ... The serious obstacle to the accomplishment of Mrs. Reid's dream was that the lack of educational background of the girls of her day made them totally unqualified to benefit by the standard of learning aimed at. By the time the College moved to 8 and 9, York Place, Baker Street, more and better girls' schools existed, so that when in 1878 the University of London opened its examinations to women, the students from Bedford College were among the first candidates. In 1881 the College recorded its first three graduates in arts, all placed in the first division, and followed the next year by the first B.Sc.; the first M.A. came in 1886. Small beginnings: but what a wealth of initiative, enthusiasm, perseverance and toil had gone into their attainment.

From *Nature* 21 May 1949.

attributed to tumour-derived erythropoietin. These factors are normally induced by oxygen starvation (hypoxia), because they increase the number of blood vessels and the oxygen-carrying capacity of the blood.

With these observations in mind, several groups have shown that cells lacking pVHL continue to produce these hypoxia-inducible factors, even when there's plenty of oxygen available⁴. That is, they act as though they are starved of oxygen, whether they are or not. For example, haemangioblastomas mainly comprise normal blood vessels responding appropriately to a local increase in angiogenic factors — factors that are secreted by a poorly understood stromal cell which has lost the function of pVHL^{9,10}. And if VEGF is expressed in the mouse brain, this is enough to cause blood vessels to proliferate and form a structure resembling a haemangioblastoma¹¹. Things seem to be more complicated in the kidney; although loss of pVHL has been seen in early, premalignant, renal cysts⁴, other mutations seem to be needed for conversion to a cancerous tumour.

How does this failure of pVHL to regulate hypoxia-inducible factors relate to its possible involvement in protein degradation? Maxwell and co-workers¹ now show that pVHL binds to two transcription factors — hypoxia-inducible factor (HIF)-1 α and HIF-2 α — and targets them for destruction. These proteins are transcription factors that regulate production of the messenger RNAs for hypoxia-inducible factors such as VEGF. The authors show that cells lacking pVHL do not degrade HIF-1 α and HIF-2 α when shifted to well-oxygenated conditions. So, the simplest model would be one in which

pVHL, through its association with elongin B, elongin C and Cul2, regulates the stability of HIF-1 α and HIF-2 α (Fig. 1). If these proteins are not degraded, the result is excessive formation of blood vessels.

A number of questions remain. For example, does pVHL bind directly or indirectly to HIF? And does the pVHL/elongin/Cul2 complex target HIF for ubiquitination? Are there other, biologically relevant targets for pVHL? It has, for instance, been implicated in formation of the extracellular matrix¹², regulation of extracellular pH¹³, and cell-cycle control¹⁴. The answers to these questions may help us to develop new ways to control angiogenesis and tumour formation. □

William G. Kaelin Jr is at the Howard Hughes Medical Institute and Department of Adult Oncology, Dana-Farber Cancer Institute, and Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA.

e-mail: william_kaelin@dfci.harvard.edu

1. Maxwell, P. H. *et al.* *Nature* **399**, 271–275 (1999).
2. Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. *Science* **284**, 455–461 (1999).
3. Kamura, T. *et al.* *Science* **284**, 657–661 (1999).
4. Kaelin, W. G. & Maher, E. *Trends Genet.* **14**, 423–426 (1998).
5. Lonergan, K. M. *et al.* *Mol. Cell. Biol.* **18**, 732–741 (1998).
6. Skowrya, D. *et al.* *Science* **284**, 662–665 (1999).
7. Ohta, T., Michel, J. J., Schottelius, A. J. & Xiaogang, Y. *Mol. Cell* **3**, 535–541 (1999).
8. Tan, P. *et al.* *Mol. Cell* **3**, 527–533 (1999).
9. Vortmeyer, A. *et al.* *Hum. Pathol.* **28**, 540–543 (1997).
10. Krieg, M., Marti, H. & Plate, P. *Blood* **92**, 3388–3393 (1998).
11. Benjamin, L. E. & Keshet, E. *Proc. Natl Acad. Sci. USA* **94**, 8761–8766 (1997).
12. Ohh, M. *et al.* *Mol. Cell* **1**, 959–968 (1998).
13. Ivanov, S. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 12596–12601 (1998).
14. Pause, A., Lee, S., Lonergan, K. M. & Klausner, R. D. *Proc. Natl Acad. Sci. USA* **95**, 993–998 (1998).

Apoptosis

Dead end for neurodegeneration?

Christian Haass

Fatal neurodegenerative diseases are characterized by the death of neurons, and one goal of biomedical research is to slow the progression of such diseases by inhibiting this process. Work published by Ona *et al.*¹ on page 263 of this issue, and by Gervais *et al.*² in *Cell*, could provide a basis for therapeutic treatment of two neurodegenerative disorders — Huntington's disease and Alzheimer's disease, respectively.

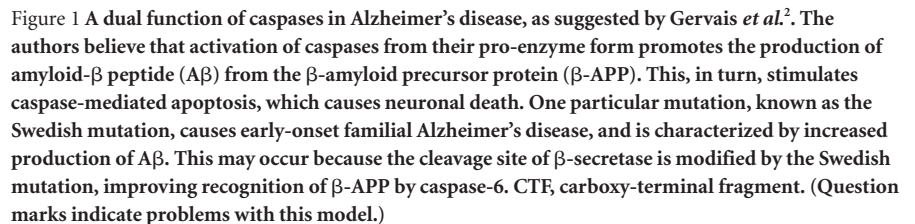
Regulated cell death (apoptosis) is not simply a pathological condition — it is critical for many normal physiological events³. Although apoptosis leads to cell loss in certain degenerative diseases, deregulation of its components can also lead to the opposite effect; uncontrolled growth of cancerous tissue. So, depending on the disease, cell death needs to be either stimulated or inhibited. In both cases, the ultimate targets for interven-

tion seem to be the key mediators of apoptosis, aspartate-specific cysteine proteases called caspases.

Like many other protein-digesting enzymes, the caspases are normally produced as inactive pro-enzymes⁴. Under certain conditions they are cleaved and activated by a process involving a cascade of such cleavages mediated by other members of the caspase family. Once active, caspases cut up their 'death substrates' by recognizing a region of primary amino-acid sequence around an essential aspartate residue at the P1 position. Cleavage of these death substrates can inactivate factors that normally protect a cell from apoptosis, destroy cellular structures, or deregulate enzymes leading to a gain or loss of function⁴.

Ona *et al.*¹ studied mice expressing the variant of the huntingtin protein (contain-

Based on these results, Gervais *et al.* have come up with a provocative model — that caspase activation promotes production of A β which, in turn, stimulates apoptosis and activates caspases (Fig. 1). Moreover, prese-



These results raise the possibility that patients with Alzheimer's disease can be treated with caspase-specific inhibitors, to slow the production of A β . But is it really so simple? Many of the results disagree with our understanding of the well-established cell biology of Alzheimer's disease, and the model created by the authors fails to accommodate previously published results. Indeed, before getting excited about a new therapeutic strategy, several discrepancies

Another thing that is not clear is how caspase-6, which is found in the cytosol, can cleave at the β -secretase site in the lumen. There is also evidence¹³ that the β -secretase cleavage of Swedish-mutant β -APP and wild-type β -APP can be blocked by serine-protease inhibitors (which would not affect caspases). Moreover, A β peptides generated from Swedish-mutant β -APP are almost exclusively cleaved on the amino-terminal side of the critical aspartate¹³, whereas the proposed caspase cleavage would occur on the carboxy-terminal side of that residue (Fig. 1).

These questions must be answered before caspase inhibition can be considered as a target for anti-Alzheimer's drugs. The road to treatment and prevention of Alzheimer's disease is, therefore, still very long. But treatment of other neurodegenerative diseases — such as Huntington's disease — may be available much sooner, because the mechanism of neuronal cell loss is better understood, and it can be mimicked in animal models. □

Christian Haass is at the Adolf Butenandt Institute, Department of Biochemistry, Ludwig-Maximilians-University, Munich 80336, Germany.

e-mail: haass@as200.zi-mannheim.de

1. Ona, V. O. *et al. Nature* **399**, 263–267 (1999).
2. Gervais, F. G. *et al. Cell* **97**, 395–406 (1999).
3. Jacobson, M. D., Weil, M. & Raff, M. *Cell* **88**, 347–354 (1997).
4. Thornberry, N. A. & Lazebnik, Y. *Science* **281**, 1312–1316 (1998).
5. Saudou, F., Finkbeiner, S., Devys, D. & Greenberg, M. E. *Cell* **95**, 55–66 (1998).
6. Davidson, F. F. & Steller, H. *Nature* **391**, 587–589 (1998).
7. De Strooper, B. *et al. Nature* **391**, 387–390 (1998).
8. Wolfe, M. *et al. Nature* **398**, 513–517 (1999).
9. Guo, Q. *et al. Nature Med.* **4**, 957–962 (1998).
10. Tanzi, R. E. *Nature Med.* **4**, 1127–1128 (1998).
11. Walter, J., Schindzielorz, A., Grünberg, J. & Haass, C. *Proc. Natl Acad. Sci. USA* **96**, 1391–1396 (1999).
12. Mattson, M. P. *Physiol. Rev.* **77**, 1081–1132 (1997).
13. Citron, M. *et al. Neuron* **17**, 171–179 (1996).

Statistical physics

Glasses go critical

Why is polystyrene brittle but polyethylene rubbery? At room temperature, the first is a glassy material whereas the second is more like a highly viscous fluid. At least, that's the easy answer. But what, exactly, is a glass? Phenomenologically, it is a liquid that has 'jammed up' — cooled without freezing until translational molecular motion is arrested. But the transition from liquid to glass poses a deep theoretical problem; indeed, the deepest and most interesting in solid-state theory, according to Philip Anderson in 1995. Elsewhere in this issue (*Nature* 399, 246–249; 1999), Sharon Glotzer and co-workers establish a link between this transition and the critical phenomena familiar from equilibrium statistical physics.

The glass transition is a transformation not between globally stable equilibrium states but between metastable states: a supercooled liquid and a disordered solid. This, then, is the fundamental difficulty — the glass transition is of a dynamical nature, representing the point at which molecules become localized and relaxation times become very long. Yet the critical phenomena most studied in statistical physics — that is, the second-order phase transition as one passes through a critical point such as the Curie point of a ferromagnet — are transitions between

Earth science

Excursions in geomagnetism

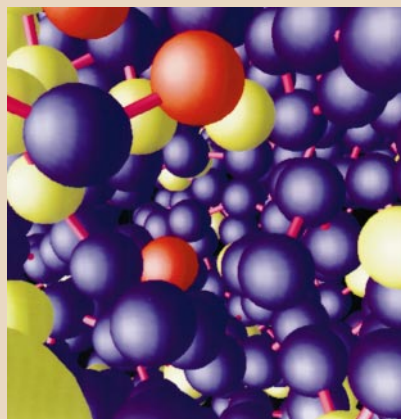
C. G. Langereis

The intensity of Earth's magnetic field varies on timescales varying from seconds to many tens of millions of years. Secular variation in the field, in terms of 100–1,000 years, can be estimated from historical records. But identifying variation on longer timescales requires continuous palaeomagnetic records from much deeper in the past, and in 1996 Guyodo and Valet¹ provided an analysis of 17 palaeointensity records going back 200,000 years. They integrated them into the so-called 'Sint-200' composite curve, which has since served as a standard. But it still covers only a comparatively brief period of Earth history.

As described on page 249 of this issue², Guyodo and Valet have now produced a new 'stack' of globally distributed records. They have analysed 33 records of relative palaeomagnetic intensity spanning the past 800,000 years, and have constructed the Sint-800 composite; the age models of all the records (which had already been normalized

with appropriate correction factors) were refined by applying the same reference curves, thereby increasing the correlation between them. The much longer time span of Sint-800 allows insight into the longer-term behaviour of the geomagnetic field — in particular, into the putative effect of climate on magnetic intensity, and into the nature and occurrence of short 'excursions' during the past 780,000 years. This is the period known as the Brunhes Chron, during which the polarity of Earth's magnetic field is as it is now; the excursions are departures from the dipole field that are considerably larger than those seen in secular variation, and sometimes even attain opposite polarity.

For some time there was speculation as to a causal relationship between Earth's magnetic field and phenomena such as climate variation and faunal extinctions. For instance when, in the 1970s, records of climatic change became available, it was proposed that change in magnetic intensity



equilibrium states, for which a static, thermodynamic description is sufficient. How to reconcile the dynamical character of the glass transition with a thermodynamic picture has recently been an issue of wide interest (see, for example, S. Sastry *et al. Nature* 393, 554–557; 1998). The key seems to be the relationship between the topographic character of the energy landscape of the system and the way in which this is explored as the system evolves from one configuration to another.

Glotzer and her colleagues take a somewhat different approach, by focusing on correlations in the atomic motions of the glass-forming medium, in this case a polymer. At a critical point, correlations between the interacting particles develop a

range much longer than the extent of the intermolecular forces. Take ferromagnetism: as the temperature is lowered through the Curie point, the spins on each magnetic particle become aligned. At the critical point itself, the correlation length of a spin becomes infinite — the spin is sensitive not just to the alignments of its near neighbours but to all others in the system.

Can such a diverging length scale be identified for the glass transition? The simulations by Glotzer and co-workers show that there seems to be a divergence in a dynamical length scale at the glass transition, namely the distance over which monomer displacements on the polymer chains are correlated. This is illustrated in the picture, a snapshot of the polymer very near the glass transition in which monomers that are moving in a correlated fashion are coloured similarly. (Red and yellow denote high mobility, and purple low mobility.) Related to this length scale is a dynamical variable that behaves like the static order parameter that characterizes equilibrium critical phenomena (the magnetization, in the case of a ferromagnet). In other words, it may be possible to mould the well-established conceptual framework of critical phenomena to the case of glass formation.

Philip Ball

could cause change in climate (nobody knew how: there was just a nice correlation). But it soon became evident that this idea stemmed from a failure to take account of variations in the amount of material magnetized, for instance through the availability of magnetic minerals which is largely under climatic control³. Because of this we now know that the intensity of the natural magnetization in sediments must be normalized by appropriate correction factors⁴.

This procedure is thought to largely remove climatic influences from the palaeomagnetic record, and the remaining traces of climatic influence are usually taken to be insignificant in perturbing the true geomagnetic signal. Supporting evidence for the geomagnetic origin of the Sint-200 composite came from a stacked record of cosmogenic ¹⁰Be in sediments⁵. This radionuclide is produced in the Earth's atmosphere, and its production depends on the degree of shielding by the Earth's magnetic field.

It has also been argued that both the Sint-200 and the ¹⁰Be stack record are not entirely free of climatic influence⁶. Nonetheless, there was almost general agreement that any trace of climate is probably caused by insufficient normalization. That consensus was abruptly shaken by a paper by Channell *et al.*⁷, published last year. Spectral analysis of two relative palaeointensity records from the North Atlantic showed apparent correlations with periodic variations in the eccentricity of Earth's orbit (100,000-year period) and in its obliquity (41,000-year period). The eccentricity component was ruled out by normalization, but not the obliquity one. Channell *et al.* therefore concluded that their record provides evidence that the Earth's obliquity influences geomagnetic field intensity, with a period of around 41,000 years. They suggested that this is caused by the effect of obliquity on precessional angular velocity and hence on precessional forces in the Earth's core.

Does this hypothesis take us back to the early 1970s? Not according to Guyodo and Valet². They performed a spectral analysis of Sint-800, comparing every 400,000-year-long interval, in steps of 100,000 years, over their entire record, and found no evidence of any dominant stable periodicity in palaeomagnetic intensity. This, however, is no doubt a debate that will run for some time.

Another feature of Sint-800 is the occurrence of many intervals of low geomagnetic intensity. Low intensities of Earth's dipole favour the incidence of geomagnetic excursions, both in direction and intensity, because the non-dipole contribution of the field becomes significant. But it is not clear whether excursions should be considered as increased secular variation or as aborted reversals; and because of the imprecision of published excursion ages, it has even been speculated that excursions correlate with climate, for example with cold periods⁸.

Guyodo and Valet² use a compilation of astronomically dated, globally occurring excursions⁹, and find that they provide an excellent match with lows in geomagnetic intensity, particularly if the field is below 50% of its present strength. Moreover, they clearly show that there is no correlation with climate, effectively killing the idea that geomagnetic excursions are related to climate.

Instead, we have a fascinating new hypothesis, formulated by Gubbins¹⁰, that relates excursions to processes in the Earth's core. The Earth's main magnetic field stems from electric currents in the liquid, iron-rich outer core. These currents derive from convective fluid flow, with a typical overturn time of about 500 years. The magnetic field of the solid inner core can change only by diffusion of the field¹¹, which has a timescale of about 3,000 years. The inner core thus stabilizes the geomagnetic field.

Gubbins¹⁰ proposes that excursions occur when the geomagnetic field reverses in the liquid outer core but not in the solid inner core. The influence of the inner core delays full reversal of the field, during which time the original polarity may establish itself in the outer core. Gubbins argues that the short duration of excursions, 10,000 years or less⁹, is consistent with the estimated diffusion time. Moreover, the large number of excursions since the last full polarity reversal 780,000 years ago, as discovered by Lund *et al.*¹² in high-resolution magnetic records from the western North Atlantic, agrees with the order-of-magnitude difference between dynamical timescales in the core.

Clearly, the Gubbins hypothesis¹⁰ requires confirmation from much longer records, while at the same time a good estimate of excursion duration requires records that are accurately dated and of high resolution.

Many excursions are not observed simply because they don't last long — more often than not, the geomagnetic signal may be distorted, or even lost, for example by post-depositional alteration of the rock after it has been magnetized. Excursions are most frequently observed during the Brunhes Chron (normal polarity, 0–780,000 years), a period that is sampled by many sedimentary cores of limited depth range. One problem with the preceding Matuyama Chron (reversed polarity, 780,000–2,580,000 years) is that it is difficult to show that normal polarity excursions are not caused by overprinting by the present-day (normal polarity) field.

So Guyodo and Valet's new, high-resolution composite of Earth's geomagnetic field tells us a great deal. From Sint-800, it now seems unlikely that geomagnetic field intensity has been influenced by Earth's orbit, or that geomagnetic excursions are related to climate; and we also have support for the new hypothesis¹⁰ that excursions and dipole fields of low intensity are an inherent property of the geodynamo. What next? We will learn a lot more on the road to Sint-2,000. □

C. G. Langereis is in the Paleomagnetic Laboratory 'Fort Hoofddijk', Faculty of Earth Sciences, Utrecht University, Budapestlaan 17, 3584 CD Utrecht, The Netherlands.

e-mail: langer@geo.uu.nl

1. Guyodo, Y. & Valet, J.-P. *Earth Planet. Sci. Lett.* **143**, 23–36 (1996).
2. Guyodo, Y. & Valet, J.-P. *Nature* **399**, 249–252 (1999).
3. Kent, D. V. & Opdyke, N. D. *Nature* **266**, 156–159 (1977).
4. Tauxe, L. *Rev. Geophys.* **31**, 319–354 (1993).
5. Frank, M. *et al.* *Earth Planet. Sci. Lett.* **149**, 121–129 (1997).
6. Kok, Y. S. *Earth Planet. Sci. Lett.* **166**, 105–119 (1999).
7. Channell, J. E. T., Hodell, D. A., McManus, J. & Lehman, B. *Nature* **394**, 464–468 (1998).
8. Worm, H.-U. *Earth Planet. Sci. Lett.* **147**, 55–67 (1997).
9. Langereis, C. G., Dekkers, M. J., de Lange, G. J., Paterne, M. & van Santvoort, P. J. M. *Geophys. J. Int.* **129**, 75–94 (1997).
10. Gubbins, D. *Geophys. J. Int.* **137**, F1–F3 (1999).
11. Hollerbach, R. & Jones, C. A. *Nature* **365**, 541–543 (1993).
12. Lund, S. P. *et al.* *EOS* **79**, 178–179 (1998).

Cell biology

Snail mail to the nucleus

Iain W. Mattaj and Elena Conti

Many macromolecules constantly move between the nucleus and cytoplasm of eukaryotic cells. Nuclear proteins and cytoplasmic RNAs, for example, must be transported from where they are made to where they act. Transit occurs via nuclear-pore complexes — large assemblies which span the nuclear envelope membranes, that separate the nucleus from the cytoplasm. Transport is mediated by receptor proteins, which shuttle between the nucleus and cytoplasm carrying cargo in one direction (Fig. 1). These receptors — known as importins and exportins — mostly belong to the importin- β protein family¹, and have also been called karyopherins. On pages 221 and 230 of this issue, Cingolani *et al.*² and Chook and Blobel³ describe the structures of

two different import receptors.

All members of the importin- β family interact with the GTP-bound form of a small GTPase called Ran. When Ran•GTP binds to importins, it causes release of the cargo (Fig. 1a); in contrast, exportins bind to their cargo only in the presence of Ran•GTP (Fig. 1b). The direction of transport depends on the concentration of Ran•GTP, which needs to be high in the nucleus and low in the cytoplasm¹. But how does binding of Ran•GTP affect the conformations of the receptors — and, so, their ability to interact with cargo? The two new receptor structures^{2,3}, one with bound cargo and the other with Ran•GTP, provide the first insights into this conformational switch.

Cingolani and colleagues² studied

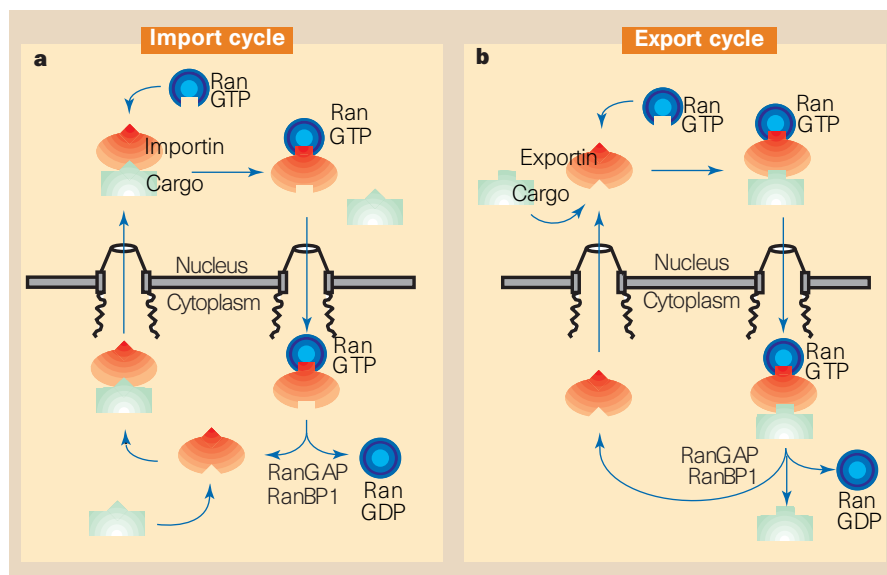


Figure 1 Nuclear import and export cycles. **a**, Import receptors bind cargo macromolecules (green) in the cytoplasm and mediate their transport to the nucleus through pores in the nuclear envelope. Binding of Ran•GTP to the importin in the nucleus causes cargo release. The Ran•GTP–receptor complex is recycled to the cytoplasm and disassembled. **b**, Export receptors act in a similar way, but the effect of Ran•GTP, in this case, is to promote (not disrupt) binding of the receptor to its cargo.

importin β , which generally cooperates with a dedicated adapter protein called importin α . This adapter binds to both the cargo molecule, and, through its 40-amino-acid importin- β -binding (IBB) domain^{4,5}, to importin β . The interaction between importin β and the IBB domain is analogous to direct receptor–cargo binding. The structure determined is importin β in its ‘cargo-bound’ conformation in complex with importin α ’s IBB domain. Chook and Blobel³ studied Karyopherin $\beta 2$ (Kap $\beta 2$; also known as transportin⁶), an importin that binds directly to a 38-amino-acid import signal⁶ called M9. These authors have solved the structure of Kap $\beta 2$ in its ‘cargo-release’ conformation in complex with Ran•GTP (Fig. 2).

The two receptors consist of protein motifs called HEAT repeats⁷ (Fig. 2a), arranged in right-handed superhelical structures. The basic HEAT repeat consists of a hairpin made up of two α -helices^{2,8} (called the A and B helices) separated by a sharp turn. A linker region connects each hairpin to the next. The A helices form the convex outer surface of the proteins, whereas the concave, inner surface of each protein is lined by B helices. This inner surface interacts directly with Ran•GTP and the IBB domain. In Kap $\beta 2$, almost all of the linkers contain a third α -helix³, but there are few linker helices in importin β ². In both receptors, one of the sharp turns is extended into a long, acidic loop, which is probably critical for Ran•GTP-mediated cargo release^{2,3}.

Most previous studies of importin β and Kap $\beta 2$ indicate that these receptors bind Ran•GTP at their amino terminus, and the target protein at their carboxy terminus. The

new results^{2,3} support this prediction. But whereas the Ran•GTP-binding regions of the two structures are similar, the substrate-binding domains are not (Fig. 2). Cingolani *et al.* find that importin β is a tightly wound superhelix, the overall structure of which resembles a snail. By contrast, Chook and Blobel show that the two halves of Kap $\beta 2$

form separate arches at roughly right angles to one another. This difference may reflect the Ran•GTP-induced conformational change, intrinsic differences between the two proteins, or both.

Why is Ran•GTP the active form of Ran? In the Kap $\beta 2$ complex, the Ran•GTP conformation is quite similar to that observed in the complex between Ran•GTP and a Ran-binding domain from the RanBP1 protein family⁹. The structure of free Ran•GDP, on the other hand, is quite different^{9,10}, and it is incompatible with receptor binding³. The RanBP1 protein family is involved in removing Ran•GTP from importins and exportins in the cytoplasm^{1,9} (Fig. 1). Consistent with this, many of the regions of contact between Ran and the Ran-binding domain⁹ are accessible in the Kap $\beta 2$ –Ran•GTP complex³ — presumably to allow Ran•GTP bound to a receptor to interact with RanBP1 and other related Ran-binding proteins *in vivo*.

Another question is how the IBB domain is bound to importin β . An α -helix in the IBB domain binds towards the carboxy terminus of importin β , whereas an extended region contacts more amino-terminal HEAT repeats, including the acidic loop (Fig. 2b). Cingolani and colleagues’ structure nicely explains the requirement for conserved basic residues in the IBB domain. Importin β ’s HEAT repeats act as a ruler, which both determines the spacing of conserved basic IBB residues and, presumably, discriminates against other positively charged potential

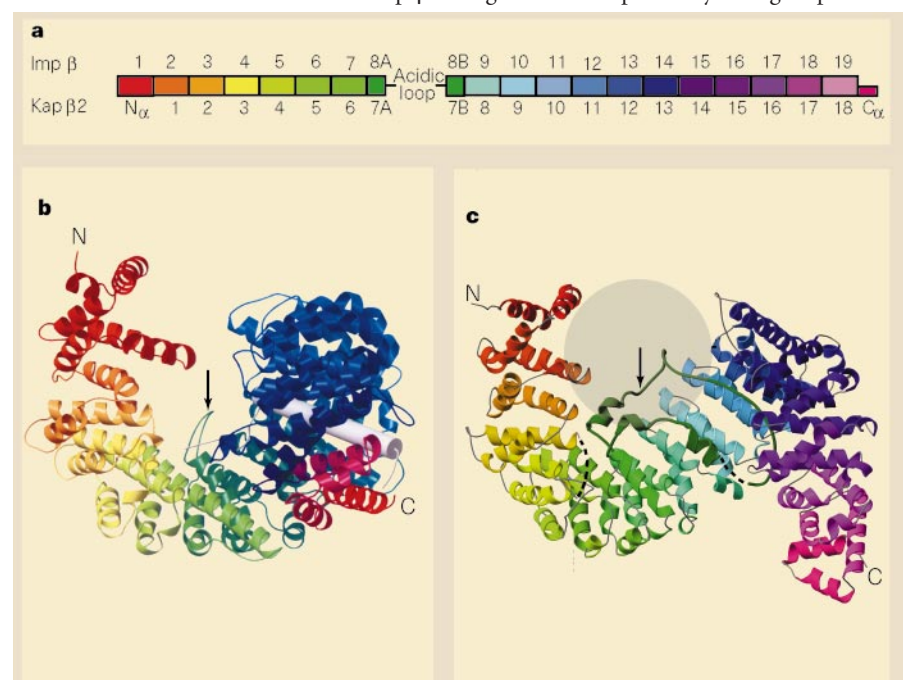


Figure 2 Comparison of importin β and transportin/Karyopherin $\beta 2$ (Kap $\beta 2$). **a**, The receptors are composed of repeated helical motifs (HEAT repeats), which are colour coded and numbered as defined^{2,3}. (Note that homologous repeats in the two receptors are numbered differently.) The structures of importin β (**b**), solved by Cingolani *et al.*², and Kap $\beta 2$ (**c**), solved by Chook and Blobel³, are both seen from the same perspective. This view illustrates the similarity between the amino-terminal halves of the receptor structures, and the dissimilarity between their carboxy-terminal halves. The acidic loops are marked by arrows. Bound importin- β -binding domain (**b**) and the Ran•GTP (**c**) are in grey.

substrates. The extensive interaction with the IBB domain may explain why the carboxy-terminal portion of importin β in the complex is very compact, and why its amino and carboxy termini are so close to one another.

How does Ran•GTP induce the importin to release its cargo? Although the acidic loops of importin β and Kap β 2 are very different^{2,3} (Fig. 2), they are probably both critical for this process. The acidic loop of importin β contacts the IBB domain directly (Fig. 2). Assuming the loop also interacts with Ran•GTP — as in Kap β 2 — then binding of Ran•GTP and the IBB domain to the loop will be mutually exclusive. One can imagine that this, plus conformational changes in the carboxy-terminal HEAT repeats owing to a steric clash with bound Ran, might progressively loosen the interactions of importin β with the IBB domain. It is harder to model the Ran switch for Kap β 2. For example, its acidic loop can be deleted without preventing binding to M9, the import signal^{6,11}, making it unlikely that mutually exclusive interactions occur in this region. Perhaps binding of Ran•GTP to Kap β 2 forces the acidic loop to cover an essential part of the M9-binding surface. Alternatively, Ran•GTP might induce a more open conformation in the carboxy-terminal arch of Kap β 2, which cannot then bind M9.

The studies of Cingolani *et al.*² and Chook and Blobel³ therefore indicate that Ran•GTP-induced structural changes in different receptors have both common and particular features. Other mysteries are deepened rather

than clarified by these important structures. First, how does binding of Ran•GTP to exportins increase — rather than disrupt — substrate binding (Fig. 1)? Second, how can an importin α molecule lacking the conserved IBB domain still be active in Ran•GTP-dependent nuclear import^{12,13}? Finally, given the limited conservation of the outer surfaces of Kap β 2 and, in particular, importin β , what mediates the specific interactions required for translocation through the nuclear pore complex? Time will tell. Luckily, the field of nucleocytoplasmic transport moves considerably faster than a snail. □

Iain W. Mattaj and Elena Conti are at the European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany.

e-mails: mattaj@embl-heidelberg.de
conti@embl-heidelberg.de

1. Mattaj, I. W. & Englmeier, L. *Annu. Rev. Biochem.* **67**, 265–306 (1998).
2. Cingolani, G., Petosa, C., Weis, K. & Müller, C. W. *Nature* **399**, 221–229 (1999).
3. Chook, Y. M. & Blobel, G. *Nature* **399**, 230–237 (1999).
4. Görlich, D., Henklein, P., Laskey, R. A. & Hartmann, E. *EMBO J.* **15**, 1810–1817 (1996).
5. Weis, K., Ryder, U. & Lamond, A. I. *EMBO J.* **15**, 1818–1825 (1996).
6. Pollard, V. W. *et al.* *Cell* **86**, 985–994 (1996).
7. Andrade, M. A. & Bork, P. *Nature Genet.* **11**, 115–116 (1995).
8. Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A. & Bradford, D. *Cell* **96**, 99–110 (1999).
9. Vetter, I. R., Nowak, C., Nishimoto, T., Kühlmann, J. & Wittinghofer, A. *Nature* **398**, 39–46 (1999).
10. Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. & Wittinghofer, A. *Nature* **374**, 378–381 (1995).
11. Fridell, R. A. *et al.* *J. Cell Sci.* **110**, 1325–1331 (1997).
12. Moroiu, J., Hijikata, M., Blobel, G. & Radu, A. *Proc. Natl Acad. Sci. USA* **92**, 6532–6536 (1995).
13. Moroiu, J., Blobel, G. & Radu, A. *Proc. Natl Acad. Sci. USA* **93**, 6572–6576 (1996).

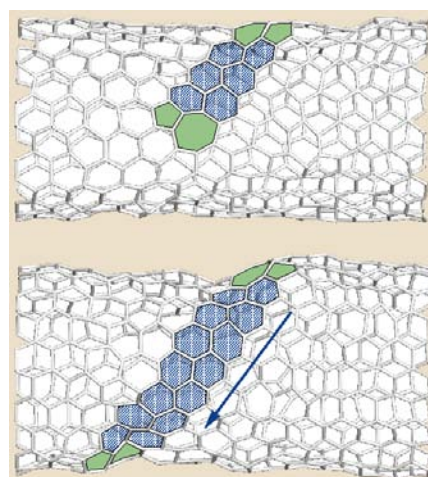


Figure 1 Plastic deformation of a carbon nanotube. Molecular dynamics simulations of a (10,10) nanotube under axial tension (J. Bernholc, M. Buongiorno Nardelli and B. Yakobson). Plastic flow behaviour is shown after 2.5 ns at $T = 3,000$ K and 3% strain. The blue area indicates the migration path (in the direction of the arrow) of the edge dislocation (green). This sort of behaviour might help make composite materials that are really tough (as measured by their ability to absorb energy).

idea that the price could drop by a factor of 10–100 if these nanotubes turn out to be really useful. They are mostly single walled, but have other forms of carbon and catalyst as impurities. Differential oxidation can be used to purify the tubes, but the oxidation opens the ends and results in 90% weight loss. In principle, the preference is to use single-walled tubes for making composites, because the inner layers of multi-walled tubes probably contribute little to carrying load, and so would reduce the stiffness for a given volume fraction of tubes. In practice, other issues need to be resolved before we can test this idea. We do have one puzzling piece of evidence. The compressive modulus of composites made with multiwalled tubes is higher than the tensile modulus¹. Textbooks tell us these moduli are supposed to be the same in normal materials — maybe it has something to do with slippage.

Now that we have nanotubes, we want to be able to disperse them individually into a polymer matrix. Unlike carbon fibres, the tubes tend to form bundles when they are produced, and it is not clear how to separate them². If the outer wall of the tube has a low surface energy like the crystal *c*-face of graphite, which is parallel to the carbon planes, polymer resins will not be able to penetrate the bundles of tubes. Water forms beads on a greasy surface for the same reason. At the meeting it was discussed, inconclusively, whether it was possible to mix the resin and tubes forcefully enough to break up the bundles.

Any nanotube reinforcement must be well bonded to the polymer matrix. We need the polymerized resin to stick to the tubes

Nanotube composites

A recipe for strength

Paul Calvert

Nanophase materials — metals or ceramics with very small grain sizes — have been fairly disappointing. It was thought that small grains should result in much harder metals. It turns out that they are harder but also more brittle. We also hoped that ceramics with very small grains would be much stronger but that hasn't really been demonstrated yet. Yet there is still reason to expect that nanometre-scale composite materials might turn out to be better than conventional composites, such as the carbon fibre–epoxy composites used in high-performance aircraft. For example, we could use nanometre-sized fibres to strengthen a polymer matrix, and carbon nanotubes are now available in sufficient quantities to test this idea. When researchers met at a conference in Alaska* to discuss the challenge of making nanocomposites, it became clear that inputs from several fields will be needed to make it work.

Carbon nanotubes should be the ideal

reinforcing fibres for composites. They have a very high aspect ratio (length-to-diameter ratio), but are short enough to flow through conventional polymer processing equipment so that parts can be moulded into complex shapes. Standard continuous-fibre composites have excellent stiffness and strength combined with low density, but are expensive to process and are limited to simple shapes, such as sheets and tubes. Short-fibre composites, on the other hand, can be moulded, but the fibres get chopped down to a maximum length of about 1 mm during the processing. At this length, the aspect ratio is only about 100 and is not enough to make a really strong material. Nanotubes can have aspect ratios of 1,000 or more, and so in theory they should make excellent composites.

Such nanotubes are indeed available as a sheet of tangled fibres (P. Eklund, Univ. Kentucky and Carbolex, Lexington, Kentucky) at prices starting from US\$100 per g. This is roughly ten times the price of gold, and so is too costly for most applications. But it is low enough to allow laboratory testing, with the

*Nanocomposites: Design and Applications Rensselaer Polytechnic Institute, Girdwood, Alaska, USA, 28 March–2 April 1999.

strongly enough so that the load is transferred to the fibres instead of the two surfaces slipping. This was a real problem in the early days of carbon-fibre composites, and oxidation was used to raise the surface energy of the fibres by exposing edges of the carbon planes. Raising the surface energy of nanotubes without destroying them will probably require some special chemistry akin to that involved in making functionalized fullerenes. Two groups (W. Blau, Trinity College, Dublin and B. Files, NASA Johnson Space Center, Houston) have made trial composites for optoelectronic and space applications, but described the mechanical properties as "quite poor". Both groups showed very long fibre pull-outs on the fractured surface of a composite, like a hair in a broken biscuit, suggesting that the interface between the resin and nanotubes was very weak.

Fullerenes are an example of promises unfulfilled. Given a totally new material, researchers envisage many possible uses ('molecular ball bearings' for lubricants being one) where fullerenes may be the best option. But from superconductors to lubricants, fullerenes have not yet improved on existing materials. Are nanotubes to be put in the same category? Probably not. Light-emitting polymer flat-panel displays are being developed, but there is a problem because the low electrical conductivity of the polymer limits the current flow. Carbon fibres or metal flakes would improve conductivity, but would also absorb light. The Trinity College work was directed at using nanotubes in electroluminescent polymers to increase electrical conductivity without detracting from the transparency³. This is also an interesting case, as specific interactions between the highly conjugated polymer and the tube were studied. The extended electron π -bonding of the polymer could provide a strong attachment to the π -bonded system of the tubes. The evidence is not clear yet, but if this was the case it might offer a method to design dispersing agents and resins to use with the tubes.

In a modern car factory, efficient spray painting of vehicles is assured by giving the paint spray an electrostatic charge, which is then attracted to the car body by the high voltage passing through it. This works fine when car bodies are made of sheet steel, but increasing use of composite materials makes it more difficult to apply an even coat. Blends of nanotubes in thermoplastics have been used for electrostatic painting of car body panels by adding enough filler to the panels to make them conducting, but not so much that they suffer from increased brittleness (T. P. Feist, General Electric, Schenectady). The high aspect ratio of the fibres means that the volume of nanotubes required is low at the percolation threshold where there are just enough fibres to form a continuous con-

ducting network through fibre-fibre contacts. Even so, at present prices the nanotube additive would cost more than the car.

Nanotubes and nanoparticles provide a simple system for molecular modellers because a tube is small enough to be treated by molecular mechanics as if it were a single, very large molecule. In macroscopic materials, a small volume can be modelled, but then boundary conditions become a concern. An intriguing possibility arises from a computer simulation that shows nanotubes collapse inwards and deform plastically when highly extended⁴ (J. Bernholc, North Carolina State Univ., Fig. 1). This type of deformation is also characteristic of tensile failure of spiralled cardboard tubes and hollow plant cells. Much of the toughness of wood can be

attributed to collapse of plant cells. If we can make these composites work, they might turn out to be really tough. Although current composites are light, stiff and strong, toughness is a worry (for instance in the limited ability of composite car bodies to absorb energy during a collision). Most new materials find applications quite unlike anything originally planned. It is just possible that nanotubes might break this rule and turn out to be good for composites after all. □

Paul Calvert is in the Department of Materials Science and Engineering, University of Arizona, Tucson, Arizona 85721, USA.

e-mail: calvert@engr.arizona.edu

1. Schädler, L. S. *et al. Appl. Phys. Lett.* **73**, 3842–3844 (1988).
2. Chen, J. *et al. Science* **282**, 95–98 (1998).
3. Coleman, J. N. *et al. Phys. Rev. B* **58**, R7492–R7495 (1998).
4. Bernholc, J. *et al. Appl. Phys. A* **67**, 39–46 (1998).

Neurobiology

Rain Man's revelations

Niels Birbaumer

In the film *Rain Man*, an autistic savant who can multiply enormous numbers with lightning speed earns social attention and wins his brother a large sum of money in a Las Vegas casino by remembering all the cards in the game. Other retarded savants, such as the three-and-a-half-year-old Nadia described by Selfe¹, draw natural objects such as the horse in Fig. 1 without any training — something that an average child and many adults cannot do. Some of these unusual people have absolute pitch; others have 'eidetic' memory, which enables them to recall more than 30 small objects presented to them for only several milliseconds. In *Proceedings of the Royal Society*, Snyder and Mitchell² now offer a theoretical explanation for these, and related, phenomena. They offer startling insights into a seemingly unusual ability, and suggest that we all may possess similar talents but do not — or cannot — use them, because we process information in a concept-driven way.

Savants are assumed to have access to low-level information processing by circumventing conscious 'executive' brain function. For most people, the route to such low-level processing may open up only during severe physical illness or altered states of consciousness. But according to Snyder and Mitchell, the savants' low-level processing is effortless and it does not require training. The skill seems to occur mainly in young savants, before the age of puberty. Maturation induces concept-driven processing of information, which requires extensive (and time-consuming) comparisons of incoming information and concepts stored in working memory.

What is the nature of these fast, low-level, elementary mental operations, and why do some — mostly pathological — conditions favour access to them? Snyder and Mitchell²



Figure 1 Drawing of a horse by a three-and-a-half-year-old autistic child (from ref. 1).

have identified one central mental operation performed by savants — the ability to partition brain activity precisely in space and time. For example, the manipulation of large numbers seems to depend on their being separated into groups and patterns. To understand the mental processes involved, we can look at the early components (up to about 100 ms) of event-related brain potentials (ERPs), which reflect what happens at the initial, 'preconscious' stages of mental processing. Later components of these ERPs involve the controlled executive functions that dominate normal information processing.

Brain activity can be precisely resolved in time and space by combining functional magnetic resonance imaging techniques, which measure local changes in blood flow, with recordings of ERPs that can track the time course of neural events. Such techniques have been used to study preconscious and conscious visual attention³. In visual attention, the conscious stages of feature

analysis and pattern recognition have been found to occur, at the earliest, 150 ms after a visual stimulus is presented. After 50 ms, the primary visual areas of the brain encode continuity or vicinity (that is, are one or several separate objects processed, and how distant are they?) and other 'local' features. At about 70–100 ms, initial preconscious attentional modulation occurs (such as early processing of 'global' features, including motion, disparity, symmetry and colour). At about 100–140 ms, the contours and parts of objects (figural primitives⁴) are extracted and pattern recognition takes place. This timescale is prolonged as the length and complexity of the stimulus increase, and a similar (but even faster) processing cascade has been identified in the auditory domain.

From these discoveries, we can deduce that savants operate on the early stages of processing — they do not progress into the later, multisensory comparisons characteristic of conscious processing. Because creativity involves the formation of such multisensory interactions in widely distributed areas of the brain, access to the low-level, early processing stages is conceptually unproductive. But Snyder and Mitchell² contend that the savants' abilities are not islands of genius in an otherwise compromised brain, and that these people do not have 'better' brains for arithmetic and other talents. Instead, the authors insist that we all have the same fundamental cognitive abilities, but that savants have access to the rapid and early low-level processing owing to a functional or pathological loss of the executive brain centres.

For example, the ERPs of the 'human calculator' in Fig. 2, a computer science student who is not autistic⁵, have the same sequence of components — and, therefore, the same stages of information processing — as those of the much slower control subjects. But looking at the amplitudes, those of the human calculator are increased compared with the controls early on, indicating enhanced automatic low-level processing. In the later stages of processing, the human calculator shows smaller amplitudes, indicating less cortical activity.

Evidence from such non-savant geniuses contradicts the findings of Snyder and Mitchell², as do results from extensive training of self-perception of low-level neuro-electric phenomena in healthy controls and neurological patients⁶. Under such training conditions, a person observes his or her own changes in slow brain potentials — indicative of high or low excitability — on a computer screen, and learns to modify it by trial-and-error learning. Snyder and Mitchell believe that training and intensive learning do not open the door to the early processing that is apparent in such astonishing performances. But the human calculator in Fig. 2 began to calculate increasingly diffi-

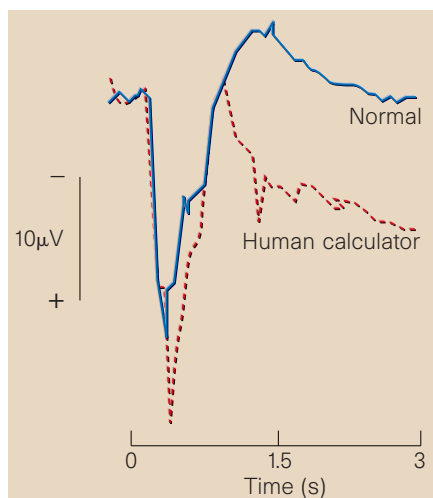


Figure 2 Event-related brain potentials (ERPs) during mental arithmetic. The solid line shows the averaged ERPs of 13 people; the dotted line, the ERP of a 'human calculator', G.M., of the same age and general intelligence as the healthy controls. G.M. solved the multiplication problems presented at point 0 (left part of the curves) twice as quickly as the controls. The problems were presented on the screen at time 0, and remained until the subject solved them (on average, about 1,100 ms for the controls and 550 ms for G.M.). The early, positive electrocortical component (left side of the curves), reflecting fast lower-level processing, is increased in the human calculator, whereas the controls show larger late, negative components (right side of the curves), indicating higher-level 'executive' processing. (Figure courtesy of P. Pauli⁵.)

cult sequences of numbers at the age of four, and continued this mental practice for more than an hour a day until he was tested 24 years later. After many training sessions, brain-damaged patients and healthy controls can learn to control their own brain activity — which we normally cannot consciously perceive — and become sensitive to the changes that underlie the elementary, low-level processing⁶. By bringing their brain activity under voluntary control, these patients can use it to communicate from the mysterious early state of non-consciousness. The result may ultimately be that we find we all possess behavioural and physiological strategies to modify the early stages of cortical processing, with unprecedented consequences for behaviour and self-awareness. □

Niels Birbaumer is at the Institute of Behavioral Neurobiology, University of Tübingen, Gartenstraße 29, 72074 Tübingen, Germany.

e-mail: niels.birbaumer@uni-tuebingen.de

1. Selfe, L. Nadia: A Case of Extraordinary Drawing Ability in Children (Academic, London, 1977).
2. Snyder, A. W. & Mitchell, D. J. *Proc. R. Soc. Lond. B* **266**, 587–592 (1999).
3. Martinez, A. et al. *Nature Neurosci.* **2**, 364–369 (1999).
4. Singer, W. in *The Mind-Brain Continuum* (eds Llinás, R. & Churchland, P. R.) 102–130 (MIT Press, Cambridge, Massachusetts, 1996).
5. Pauli, P. et al. *Psychophysiology* **33**, 522–529 (1996).
6. Birbaumer, N. et al. *Nature* **398**, 297–298 (1999).

Daedalus

Shear chemistry

On the molecular level, cutting is a very brutal process. The cutting tool forces a crack through the material, creating two highly reactive surfaces. Indeed, some quite soft materials, such as rubber and ebonite, rapidly blunt the hardest cutter. The ruptured chemical bonds of the new surfaces, dense with free radicals, combine with the cutter and erode it rapidly. Daedalus wants to improve the fundamentals.

The most efficient way to cut a material, he says, is to raise the molecules along the line of the cut into an anti-bonding excited state. The workpiece would not only lose all strength along the line of the cut; the two sides would actually repel each other. A laser beam of the right frequency should do the trick. It need not even be very intense. At any instant, it merely has to excite a single line of molecules. His prototype cutter has a stack of optic fibres which project the light in a narrow parallel beam out from its leading edge.

By itself, this simple concept would never work. As soon as the separated surfaces passed out of the micrometre-wide laser beam, they would relax back to their ground state — two dense sheets of free radicals sliding back over either side of the cutter. They would react with it aggressively. Behind its leading edge, the anti-bonding cutter will need to exude a reactive cutting lubricant, to combine with and stabilize the newly created surfaces, and give them the lowest possible friction. Fluoro-polymers have very low friction, so Daedalus is exploring cutting lubricants containing easily decomposed fluorides, such as those with xenon. Again, because only a monolayer of molecules need be supplied, very little of the expensive reagent will be needed.

Daedalus's anti-bonding cutter will be a costly, sophisticated device. Its laser wavelength and chemical cutting lubricant will need to be specially chosen for each material it has to cut. But it will then glide through the hardest and most refractory materials with wonderful ease and speed, giving optically smooth surfaces. Mated with that other sophisticated advance, the numerically controlled lathe or milling machine, it will bring new power to industry. To fund the venture, however, Daedalus is seeking military support. An anti-bonding armour-piercing smart shell or bomb would penetrate the most hardened defences as if they were not there.

David Jones

Birds extend their ranges northwards

We have analysed the breeding distributions of British birds over a 20-year period. After controlling for overall population expansions and retractions, we find that the northern margins of many species have moved further north by an average of 18.9 km during this time. This general northward shift took place during a period of climatic warming, which we propose might be causally involved.

Butterfly species are moving northwards in Europe¹, probably in response to climate warming^{2,3}. Studies of other species that show range expansions at one margin closer to the pole or towards a higher elevation^{4–6} can be difficult to interpret because changes at single margins might be part of overall population expansions, without any true bias from elevation or latitude^{7,8}. Here we control for this effect, using data on British birds, and still find that northern margins have expanded northwards.

We used data from two atlases of breeding birds compiled for Great Britain, based on a 100-km² grid and encompassing the periods 1968–72 (ref. 9) and 1988–91 (ref. 10). Overall increases in distribution size, whatever the cause, would be expected to cause species restricted to the south of Britain (for example, *Cettia cetti*)¹⁰ to move northwards at their northernmost boundary, and species restricted to the north (such as *Mergus merganser*)¹⁰ to move southwards. In contrast, a general decline in a species

would be expected to cause northern species (*Tetrao tetrix*, for instance)¹⁰ to retract northwards, and southern species (such as *Streptopelia turtur*)¹⁰ to retract southwards, towards their cores⁸. These patterns arise because the most marginal records are likely to be further from the distributional core in a period with more 100-km²-grid records of a species. They do not reflect any systematic shift north or south.

For terrestrial or freshwater bird species with southerly distributions in Britain, we recorded the mean location of the ten most northerly grid squares with breeding in 1968–72 (ref. 9) and 1988–91 (ref. 10). As expected, the northern margins of most southerly species that increased in numbers of squares over the 20-year period shifted northwards, and the northern margins of many southerly species that declined overall shifted southwards (Fig. 1a). The $x=0$ intercept in Fig. 1a shows the mean shift in range for species with no overall change in the number of grid squares occupied: a northward shift of 18.9 km ($P=0.0054$). The left-hand point in Fig. 1a could have affected the result, so we re-analysed the relation by omitting this point and then leaving out the two left-hand points. These produced intercept values of 18.3 ($P=0.0064$) and 19.4 ($P=0.0033$) km shifts to the north.

For northerly species, we recorded the mean locations of the ten southernmost

grid squares with breeding in 1968–72 (ref. 9) and 1988–91 (ref. 10). The southern margins of northerly species that increased in numbers of squares over the 20-year period generally shifted southwards, whereas the southern margins of most species that declined shifted northwards (Fig. 1b). For these species, there has been no systematic shift northwards or southwards (intercept, 4.5 km south; $P=0.62$).

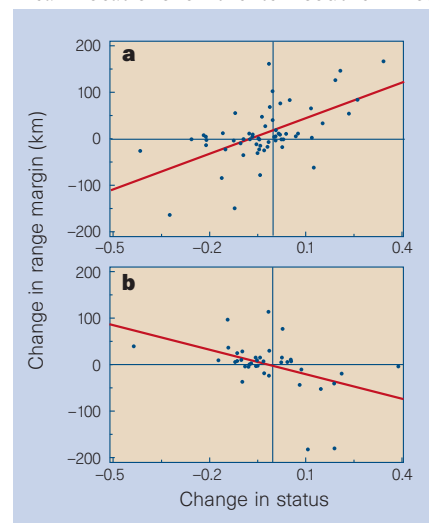
Many factors might have caused the distributions of particular species to change in individualistic ways. Nevertheless, after taking account of changes in the numbers of grid squares occupied, we estimate that the northern margins of southerly species have moved north by an average of 18.9 km in 20 years. The most parsimonious explanation is climate, given that: (1) recent changes in the timing¹¹ and success¹² of bird reproduction are associated with warmer spring conditions; (2) the overall spatial distribution of bird diversity in Britain is correlated with that of temperature¹³; (3) summer temperature was a significant predictor of the breeding distribution of 87 of 194 (45%) terrestrial and freshwater bird species (J. J. L., J. J. D. Greenwood and J. R. G. Turner, unpublished observations); and (4) the shifts coincided with a period of climate warming^{2,3}. In contrast, the southern margins of northerly species have not shifted consistently. This parallels the result obtained for European butterflies, for which the northern margins have expanded more than the southern margins have retracted¹. Cool margins of temperate species might be more immediately responsive than warm margins to the direct effects of thermal variation; warm margins might respond more to rainfall, species interactions or longer-term climate-related changes in the vegetation.

Chris D. Thomas, Jack J. Lennon

Centre for Biodiversity and Conservation,
School of Biology, University of Leeds,
Leeds LS2 9JT, UK

e-mail: c.d.thomas@leeds.ac.uk

Figure 1 Change in the range margins of bird species (defined as the mean distance northwards of the ten most marginal records in 1988–91 minus the mean of the ten most marginal records in 1968–72; positive values indicate a move northwards, and negative values a move southwards) plotted against changes in status ($\log_{10}$ of the numbers of 100-km² grid cells occupied in 1988–91 compared with 1968–72). Only 100-km² grid cells mapped in both atlases were analysed. **a**, Plot for 59 southerly species for which the mean location of the occupied 100 km² was south of the mean location of all 100-km² cells in Great Britain; the northern margin was analysed. The equation for the regression line is $x=18.87$ km (s.e.m. 6.52, $P<0.0054$) + 255.63 (s.e.m. 47.47, $P<0.0001$); $r^2=0.337$. **b**, Plot for 42 northerly species for which the mean location of the occupied 100 km² was north of the mean location of all 100-km² cells; the southern margin was analysed. The equation of the regression line is $x=-4.48$ km (s.e.m. 8.96, $P=0.6200$) - 183.45 (s.e.m. 43.07, $P<0.0001$); $r^2=0.312$. The following categories of species were excluded: those occupying <20 cells, to avoid counting southernmost as well as northernmost records for the rarest species; ubiquitous species occupying >2,000 out of the 2,280 cells available; recently introduced species; species for which grid cells were moved or omitted on published maps. A larger sample of 95 southerly species that included species occupying 2,000–2,280 grid cells was also significant (intercept, 13.33 km (northwards); $P=0.0016$), but this underestimates the true shift because species occupying 2,000–2,280 cells had less opportunity to colonize new grid cells northwards. The result for northerly species was unaffected by the inclusion of two extra species occupying 2,000–2,280 cells (intercept, -3.90 km (southwards), $P=0.6505$).



1. Parmesan, C. *et al.* *Nature* (in the press).
2. Easterling, D. R. *et al.* *Science* **277**, 364–367 (1997).
3. Conway, D. *Prog. Phys. Geogr.* **22**, 350–374 (1998).
4. Taylor, R. H. & Wilson, P. R. N. Z. J. *Ecol.* **14**, 25–29 (1990).
5. Grabherr, G., Gottfried, M. & Pauli, H. *Nature* **369**, 448 (1994).
6. Epstein, P. R. *et al.* *Bull. Am. Meteorol. Soc.* **79**, 409–417 (1998).
7. Parmesan, C. *Nature* **382**, 765–766 (1996).
8. Melham, D. W. *Ecol. Appl.* **7**, 614–624 (1997).
9. Sharrock, J. T. R. *The Atlas of Breeding Birds in Britain and Ireland* (Poyser, Berkhamsted, 1976).
10. Gibbons, D. W., Reid, J. B. & Chapman, R. A. *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991* (Poyser, London, 1993).
11. Crick, H. Q. P., Dudley, C., Glue, D. E. & Thomson, D. L. *Nature* **388**, 526 (1997).
12. Visser, M. E., van Noordwijk, A. J., Tinbergen, J. M. & Lessells, C. M. *Proc. R. Soc. Lond. B* **265**, 1867–1870 (1998).
13. Turner, J. R. G., Lennon, J. J. & Lawrenson, J. A. *Nature* **335**, 539–541 (1988).

Transgenic pollen harms monarch larvae

Although plants transformed with genetic material from the bacterium *Bacillus thuringiensis* (*Bt*) are generally thought to have negligible impact on non-target organisms¹, *Bt* corn plants might represent a risk because most hybrids express the *Bt* toxin in pollen², and corn pollen is dispersed over at least 60 metres by wind³. Corn pollen is deposited on other plants near corn fields and can be ingested by the non-target organisms that consume these plants. In a laboratory assay we found that larvae of the monarch butterfly, *Danaus plexippus*, reared on milkweed leaves dusted with pollen from *Bt* corn, ate less, grew more slowly and suffered higher mortality than larvae reared on leaves dusted with untransformed corn pollen or on leaves without pollen.

Pollen for our assay was collected from N4640-*Bt* corn and an unrelated, untransformed hybrid, and was applied by gently tapping a spatula of pollen over milkweed (*Asclepias curassavica*) leaves that had been lightly misted with water. Pollen density was set to visually match densities on milkweed leaves collected from corn fields. Petioles of individual leaves were placed in water-filled tubes that were taped into plastic boxes. Five three-day-old monarch larvae from our captive colony were placed on each leaf, and each treatment was replicated five times. Milkweed leaf consumption, monarch larval survival and final larval weight were recorded over four days.

Larval survival (56%) after four days of feeding on leaves dusted with *Bt* pollen was significantly lower than survival either on leaves dusted with untransformed pollen or on control leaves with no pollen (both 100%, $P=0.008$) (Fig. 1a). Because there was no mortality on leaves dusted with untransformed pollen, all of the mortality on leaves dusted with *Bt* pollen seems to be due to the effects of the *Bt* toxin.

There was a significant effect of corn pollen on monarch feeding behaviour ($P=0.0001$) (Fig. 1b). The mean cumulative proportion of leaves consumed per larva was significantly lower on leaves dusted with *Bt* pollen (0.57 ± 0.14 , $P=0.001$) and on leaves dusted with untransformed pollen (1.12 ± 0.09 , $P=0.007$) compared with consumption on control leaves without pollen (1.61 ± 0.09). The reduced rates of larval feeding on pollen-dusted leaves might represent a gustatory response of this highly specific herbivore to the presence of a 'non-host' stimulus. However, such a putative feeding deterrence alone could not explain the nearly twofold decrease in

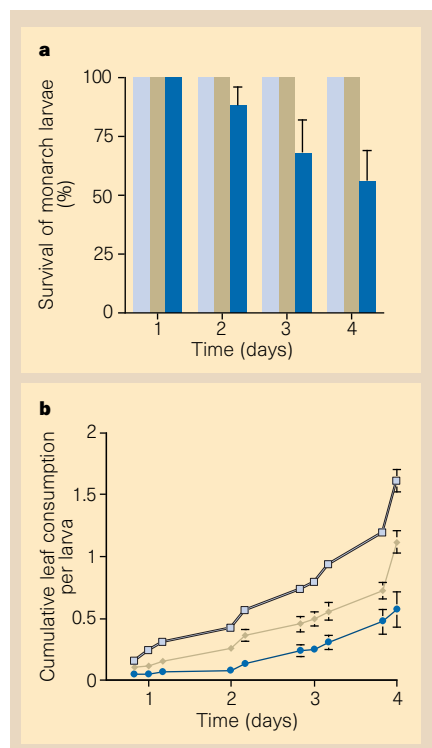


Figure 1 Survival and leaf consumption of second- to third-instar monarch larvae on each of three milkweed leaf treatments: leaves with no pollen (light blue), leaves treated with untransformed corn pollen (green) and leaves dusted with pollen from *Bt* corn (dark blue). **a**, Mean ($\pm$ s.e.m.) survival based on the proportion of larvae surviving in five replicates of each treatment. **b**, Mean ($\pm$ s.e.m.) cumulative leaf consumption based on the total amount of leaf area consumed per larva in five replicates of each treatment. The amount of leaf area consumed per larva in each experimental unit was calculated for each time interval by dividing the amount of leaf area consumed in that interval by the number of larvae alive during the time interval. Cumulative consumption was calculated by summing the leaf area consumed per larva at each interval. Colours of lines correspond to those of the bars in **a**.

consumption rate on leaves with *Bt* pollen compared with leaves with untransformed pollen ($P=0.004$).

The low consumption rates of larvae fed on leaves with *Bt* pollen led to slower growth rates: the average weight of larvae that survived to the end of the experiment on *Bt*-pollen leaves (0.16 ± 0.03 g) was less than half the average final weight of larvae that fed on leaves with no pollen (0.38 ± 0.02 g, $P=0.0001$).

These results have potentially profound implications for the conservation of monarch butterflies. Monarch larvae feed exclusively on milkweed leaves⁴; the common milkweed, *A. syriaca*, is the primary host plant of monarch butterflies in the northern United States and southern Canada⁵. Milkweed frequently occurs in and around the edges of corn fields, where it is fed on by monarch larvae⁶. Corn fields

shed pollen for 8–10 days between late June and mid-August, which is during the time when monarch larvae are feeding⁷. Although the northern range of monarchs is vast, 50% of the summer monarch population is concentrated within the mid-western United States, a region referred to as the 'corn belt' because of the intensity of field corn production⁸. The large land area covered by corn in this region suggests that a substantial portion of available milkweeds may be within range of corn pollen deposition.

With the amount of *Bt* corn planted in the United States projected to increase markedly over the next few years⁹, it is imperative that we gather the data necessary to evaluate the risks associated with this new agrotechnology and to compare these risks with those posed by pesticides and other pest-control tactics.

John E. Losey, Linda S. Rayor, Maureen E. Carter

Department of Entomology, Comstock Hall, Cornell University, Ithaca, New York 14853, USA
e-mail: jel27@cornell.edu

- Ostlie, K. R., Hutchison, W. D. & Hellmich, R. L. *Bt Corn and European Corn Borer* (NCR publ. 602, Univ. of Minnesota, St Paul, 1997).
- Fearing, P. L., Brown, D., Vlachos, D., Meghji, M. & Privalle, L. *Mol. Breed.* **3**, 169–176 (1997).
- Raynor, G. S., Ogden, E. C. & Hayes, J. V. *Agron. J.* **64**, 420–427 (1972).
- Malcolm, S. B., Cockrell, B. J. & Brower, L. P. in *Biology and Conservation of the Monarch Butterfly* (eds Malcolm, S. B. & Zalucki, M. P.) 253–267 (Natural History Museum of Los Angeles County, Los Angeles, 1993).
- Malcolm, S. B., Cockrell, B. J. & Brower, L. P. *J. Chem. Ecol.* **15**, 819–853 (1989).
- Yenish, J. P., Fry, T. A., Durgan, B. R. & Wyse, D. L. *Weed Sci.* **45**, 44–53 (1997).
- Brower, L. P. *J. Exp. Biol.* **199**, 93–103 (1996).
- Wassenaar, L. I. & Hobson, K. A. *Proc. Natl Acad. Sci. USA* **95**, 15436–15439 (1998).
- Andow, D. A. & Hutchison, W. D. in *Now or Never: Serious New Plans to Save a Natural Pest Control* (eds Mellon, M. & Rissler, J.) 19–65 (Union of Concerned Scientists, Cambridge, Massachusetts, 1998).

The mystery of female beauty

Yu and Shepard¹ have reported a preference for heavy women with high waist-to-hip ratios (WHR) in a culturally isolated population in southeast Peru. Their findings are interesting because a preference for low WHR is widespread in westernized populations^{2–5}. However, we disagree with their argument that cultural invariance is necessary for an adaptationist interpretation of WHR preference.

WHR and waist circumference are positively correlated with testosterone and negatively associated with oestrogen⁶. Women with low WHR have better health and fertility than women with high WHR⁵. However, women in England and Texas with high

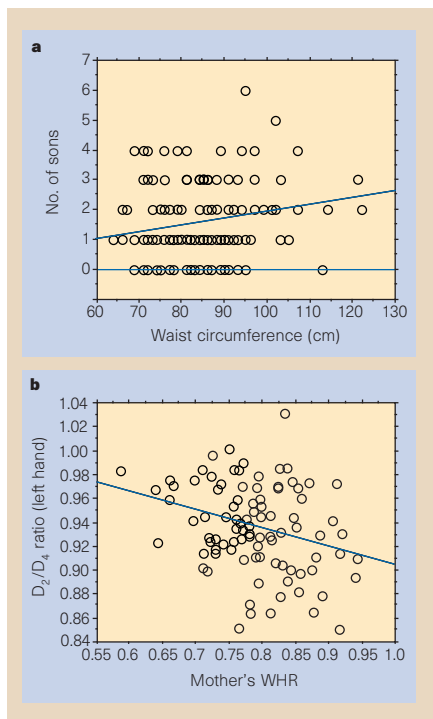


Figure 1 Relations between female body shape and the number of sons and testosterone levels in their children. **a**, Number of sons and waist circumference in 141 Jamaican women; **b**, D_2/D_4 ratio for the left hands of 95 Jamaican children and the WHR of their mothers. The lines are least-squares lines of best fit: **a**, $y = 0.023x - 0.34$; **b**, $y = -0.155x + 1.059$.

WHR and thick waists tend to have more sons^{7,8}, and a preference for women with high WHR might result in selection for increased testosterone levels in children. Men in societies that prize sons over daughters and in which strength is an advantage might therefore be expected to show a preference for high WHR.

We present data from a rural Jamaican population showing that: (1) in agreement with the data from England and Texas^{7,8}, there is a positive association between a woman's waist circumference and her number of sons; and (2) a high WHR in women is associated with a marker for high testosterone (a low ratio of the lengths of the second and fourth digits, or D_2/D_4 ; ref. 9) in their children.

Our Jamaican sample was drawn from Southfield in the parish of St Elizabeth and was part of a large long-term study of developmental stability (the Jamaican Symmetry Project). We measured 141 women whose

mean age was 34.63 ± 7.23 s.d. years. The waist was measured at the narrowest portion between the ribs and the iliac crest, and hips were measured at the level of the greatest protrusion of the buttocks. All subjects were standing and wore light outdoor clothing.

Photocopies of children's hands were used to measure digit length (95 subjects, mean age 8.00 ± 1.43 s.d. years, were asked to place hands palm-down on the glass platen). Digits were measured from the crease proximal to the palm to the tip of the digit⁹. Repeated measurements on the photocopies and comparisons with X-rays of the hands indicated that there was a high concordance with the photocopied measurements. The mean D_2/D_4 ratio for the right hand was 0.93 ± 0.03 , and 0.94 ± 0.04 for the left.

The waist circumference of mothers was significantly and positively related to the number of sons (Fig. 1a) but not to the number of daughters. The relation between a woman's waist circumference and the number of her sons remained significant after the effect of the age of the mother was partialled out (partial correlation, number of sons and waist circumference: $r = 0.18$, $P = 0.03$; and age of mother: $r = 0.30$, $P = 0.0005$).

Similar, but weaker, relationships were present between hip circumference and WHR and the number of sons. Waist and hip circumference and WHR were positively but not significantly related to the proportion of sons (Table 1). The D_2/D_4 ratio was significantly and negatively related to WHR in the right and left hands; that is, women with high WHR tended to have children with low D_2/D_4 ratios (Table 1 and Fig. 1b). As a low D_2/D_4 ratio is indicative of a high testosterone level in adults, this probably means that women with high waist-to-hip ratios have children with high testosterone.

Of course, our Jamaican sample is westernized and we have no data on WHR preferences in adults. However, the positive relation between waist circumference and/or WHR and a tendency to produce sons has now been demonstrated in three populations and may well be universal. The high testosterone found in the children of high-WHR women might relate to muscular strength (it is already known that high-WHR women have heavier and taller

babies than low-WHR women¹⁰). This might explain why traditional societies in which sons are valued over daughters may prefer tubular women, which is consistent with an adaptionist interpretation of these preferences.

J. T. Manning*, **R. L. Trivers†**, **D. Singh‡**, **R. Thornhill§**

*School of Biological Sciences, University of Liverpool, Liverpool L69 3BX, UK
e-mail: jtmann@liv.ac.uk

†Department of Anthropology, Rutgers University, New Brunswick, New Jersey 08901-1414, USA

‡Department of Psychology, University of Texas, Austin, Texas 78712, USA

§Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131-1091, USA

1. Yu, D. W. & Shepard, G. H. *Nature* **396**, 321–322 (1998).
2. Singh, D. *J. Pers. Soc. Psychol.* **65**, 293–307 (1993).
3. Singh, D. *Hum. Nat.* **6**, 51–68 (1995).
4. Furnham, A., Tan, T. & McManus, C. *Per. Indiv. Diff.* **22**, 539–549 (1997).
5. Henss, R. *Per. Indiv. Diff.* **19**, 479–488 (1995).
6. Evans, D. J., Hoffmann, R. G., Kalkhoff, R. K. & Kissebah, A. H. *J. Clin. Endocrinol. Metab.* **57**, 304–310 (1983).
7. Manning, J. T., Anderton, R. & Washington, S. M. *J. Hum. Evol.* **31**, 41–47 (1996).
8. Singh, D. & Zambardo, R. *J. Hum. Biol.* **69**, 545–556 (1997).
9. Manning, J. T., Scutt, D., Wilson, J. & Lewis-Jones, D. I. *Hum. Reprod.* **13**, 3000–3004 (1998).
10. Brown, J. E. *et al. Epidemiology* **7**, 62–66 (1996).

Evolutionary psychology suggests that a woman's sexual attractiveness might be based on cues of reproductive potential. It has been proposed that a major determinant of physical attractiveness is the ratio between her waist and hip measurements (the waist-to-hip ratio, or WHR): for example, a woman with a curvaceous body and a WHR of 0.7 is considered to be optimally attractive^{1–3}, presumably because this WHR is the result of a fat distribution that maximizes reproductive potential⁴. It follows that the preference for a curvaceous body shape in women should be universal among men and not be culturally based, because natural selection presumably favours cues indicative of the most fertile body shape.

Yu and Shepard have challenged this hypothesis⁵. They tested the preferences of a culturally isolated tribe of Peruvian Indians (the Matsigenka) by using a set of line-drawn figures of women who varied in apparent body-mass index (BMI) and WHR¹. They claim that their results indicate a preference by this tribe for a tubular body shape, rather than the curvaceous shape favoured in the United States⁵. However, we believe that the conclusions of Yu and Shepard are undermined by a flawed assumption.

The drawings used by Yu and Shepard are arranged in three series⁵: underweight, normal and overweight. Within each series, the BMI of each of the four figures is supposed to be held constant, while the WHR is varied by narrowing the waist. However, we believe that this assumption is false

Table 1 Effects on offspring of WHC and WHR of Jamaican women

Trait	No. of sons		No. of daughters		Proportion of sons		D_2/D_4 right hand		D_2/D_4 left hand	
	r	P	r	P	r	P	r	P	r	P
Waist circumference	0.21	0.01	0.02	0.79	0.17	0.052	–0.15	0.15	–0.18	0.08
Hip circumference	0.14	0.09	0.07	0.41	0.16	0.06	–0.02	0.87	–0.01	0.92
WHR	0.15	0.08	0.05	0.56	0.10	0.27	–0.20	0.048	–0.29	0.004

The relations are shown between measurements of waist and hip circumference (WHC) and waist-to-hip ratio (WHR) for Jamaican women, and the proportion of sons (arcsine transformed, $n = 141$ women) and the number of sons and daughters ($n = 141$ women) they have, and the D_2/D_4 ratio in the left and right hands of their children ($n = 95$ women). r , Coefficient of correlation; P , probability of wrongly rejecting the null hypothesis.

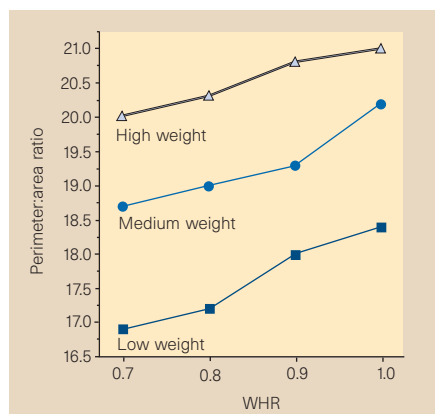


Figure 1 Correlation between perimeter-to-area ratio and WHR of figures produced by Singh¹ and used by Yu and Shepard⁵. In these line-drawn images, the BMI of the figures within each group of underweight, normal, and overweight women was assumed to be constant while their WHR was varied by narrowing the waist. However, this procedure alters not only the WHR, but also the apparent BMI: as the WHR rises, so does the apparent BMI. This correlation can be quantified by plotting the perimeter-to-area ratio of each of the line drawings used by Yu and Shepard against the WHR. The two measures are highly correlated ($r \geq 0.95$ for each of the three series).

because, when the figures are modified by altering the width of the torso around the waist, this alters not only the WHR but also the apparent BMI: as the value of the WHR rises, so does the apparent BMI^{6,7}.

This co-variation can be quantified. Using photographs of real women (with known BMI values), we have measured the path length around the perimeter of each figure and divided it by the area within the perimeter (the perimeter-to-area ratio). This ratio correlates well with the actual BMI of the woman (Pearson correlation coefficient is $r = 0.97$, $P < 0.0001$). We can then calculate the perimeter-to-area ratio for each of the drawings used by Yu and Shepard, and, by plotting these values against WHR, we can show that the two measures are clearly correlated ($r \geq 0.95$ for each of the three series) (Fig. 1).

This co-variation of BMI with WHR in the line-drawn figures means that if a subject chooses the least curvaceous shape, he is also choosing the highest BMI. BMI is a very strong cue to attractiveness^{6,7}, and this sample of Matsigenka people seem to prefer women with a high BMI (that is, the figures in the 'overweight' set). So it is quite possible that their reported preference for a less curvaceous shape merely represents a preference for a figure with a high BMI. Yu and Shepard's data set cannot differentiate between a preference for a higher BMI and/or a less curvaceous figure. Firm conclusions about the Matsigenka's preferences for female physical shape need to be based on a set of stimuli in which BMI and WHR

vary independently, so that the relative importance of the two features can be determined.

M. J. Tovée, P. L. Cornelissen

Department of Psychology, Ridley Building,
Newcastle University,

Newcastle Upon Tyne NE1 7RU, UK

e-mail: m.j.tovee@ncl.ac.uk

1. Singh, D. *Hum. Nat.* **4**, 297–321 (1993).
2. Henss, R. *Pers. Individ. Diff.* **19**, 479–488 (1995).
3. Tovée, M. J., Mason, S. M., Emery, J. L., McCluskey, S. E. & Cohen-Tovée, E. M. *Lancet* **350**, 1474–1475 (1997).
4. Zaadstra, B. M. *et al. Br. Med. J.* **306**, 484–487 (1993).
5. Yu, D. W. & Shepard, G. H. *Nature* **396**, 321–322 (1998).
6. Tovée, M. J., Reinhardt, S., Emery, J. L. & Cornelissen, P. L. *Lancet* **352**, 548 (1998).
7. Tovée, M. J., Maisey, D. S., Emery, J. L. & Cornelissen, P. L. *Proc. R. Soc. Lond. B* **266**, 211–218 (1999).

Yu and Shepard reply — We have proposed that cultural invariance in beauty preferences could be an artefact of exposure to a dominant culture, and also that evolutionary psychology should embrace variation because adaptive evolution is as likely to produce variable outcomes as fixed ones¹. Accordingly, Manning *et al.* suggest that a culturally variable male preference for women with high WHR might reflect a culturally variable preference for producing more sons with higher testosterone levels. But this intriguing explanation probably does not apply to the Matsigenka people.

The Matsigenka marriage system is matrilineal, meaning that a man leaves his natal home to marry into the woman's family. As a result, families lose their economic investment in their sons, not daughters, and so sons are more costly than daughters. Also, elder sons traditionally marry polygamously, meaning that the expected reproductive success of younger sons is lower than that of any daughter, and thus, if anything, there should be a preference for female offspring. As it turns out, sex ratios seem to be female-biased. A genealogical survey of an even more culturally isolated group of Matsigenka, the Sotileja population, recorded a birth ratio of 73 females to 62 males, although this female bias was later reduced by higher female mortality².

Nor does the model of Manning *et al.* explain why westernized indigenous populations in South America should prefer hourglass-shaped females. First, what traditional society does not value strength (or sons)? Second, Peru's male-dominated economy should increase, not decrease, the value of males in westernized populations, and, by extension, the value of high-WHR females. Finally, how could such preferences change evolutionarily in a single generation?

Tovée and Cornelissen suggest that WHR is a proxy measurement for BMI. In Yombyato village, under all three decision criteria, figures were ranked O9, O7, N9, N7, U9, U7 (sorted by median then mean ranks; notation: O, N, U are overweight, normal and underweight, respectively, and numbers refer

to high (9) and low (7) WHRs), which is consistent with the hypothesis that figures were ordered by BMI alone. However, Shipetiari males ranked the figures O7, O9, N7, N9, U7, U9 under the attractiveness and spouse criteria, which might not be consistent with the BMI hypothesis, because the ranks do not vary smoothly with BMI. Moreover, Alto Madre and United States males generally grouped figures first by WHR (7, 9) and then by weight, which is inconsistent with the BMI hypothesis. Also, although both the latter populations generally ranked N7 first, O7 was the second choice of Alto Madre males, and U7 was that of United States males. On the other hand, by using a different set of (headless) figures, Tovée *et al.* found that British university students, in contrast, strongly disliked females with low or high values of BMI but were agnostic with regard to WHR³. Finally, a recent study has revealed that the culturally isolated hunter-gatherer Hadza men in Tanzania are also indifferent to WHR but prefer pictures of high-BMI women (A. Wetsman and F. Marlowe, personal communication).

What can we make of these cross-cultural differences? First, not only might WHR preferences be affected by westernization, but the very concept of using WHR at all to assess attractiveness might be culturally variable and, in Peru, an effect of westernization. Second, these results show that BMI preferences vary markedly across cultures (as well as through time⁴). If we were to rely on the weak cultural invariance test of adaptation, BMI and WHR preferences would not be considered as adaptations.

We obviously need a more sophisticated evolutionary analysis to explain variation in preferred body shapes and sizes, which does not ignore the effects of environment on ontogeny. For now, the nature of female beauty must remain mysterious, as perhaps it should.

Douglas W. Yu*, Glenn H. Shepard†

*Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA

(e-mail: d.yu@ic.ac.uk)

†Department of Anthropology, University of California, Berkeley, Berkeley, California 94720 USA

1. Widemo, F. & Saether, S. A. *Trends Ecol. Evol.* **14**, 26–31 (1999).
2. Cueva, N. & Shepard, G. H. Informe sobre la visita de las comunidades Machiguenga en el río Sotileja. (Asociación para la Conservación de la Naturaleza (APECO), Lima, Peru, 1995).
3. Tovée, M. J., Reinhardt, S., Emery, J. L. & Cornelissen, P. L. *Lancet* **352**, 548 (1998).
4. Singh, D. *J. Pers. Soc. Psychol.* **65**, 293–307 (1993).
5. Yu, D. W. & Shepard, G. H. *Nature* **396**, 321–322 (1998).

erratum

Colour was inappropriately added by *Nature* to Fig. 1 of Yu and Shepard's original paper⁵, which should have appeared in black and white. We apologize for any confusion that may have arisen as a result.

Looping the evolutionary loop

The Origins of Life: From the Birth of Life to the Origin of Language

by John Maynard Smith and
Eörs Szathmáry

Oxford University Press: 1999. 180 pp. £18.99

Gabby Dover

One of the sadistic pleasures to be had from the defunct age of selfish-genery was to witness the mental loops of its proponents as they tried to extricate themselves from the illogical cul-de-sacs of their own devising. In his writings, Richard Dawkins' pseudo-"paradox of the organism" was the climactic apotheosis of a belief in his own rhetorical devices which forced him to suspend all scientific rationale and modesty.

The argument is as follows and, in the interests of fairness, I quote directly: "the organism should, by rights, be torn apart by its competing replicators", yet "the organism functions as such a convincingly unified whole that biologists in general have not seen there is a paradox at all!". The genes get around this little problem by deciding on a minimum shared list of "desiderata" of what to do: "They all 'agree' over what is the optimum state of every aspect of phenotype, all agree on the correct wing length, leg colour ... etc.". This Dawkinsesque loop now begs the question: which is the unit of selection? Is it the selfish gene or the organism (the collective love-in of happy, hippy desiderata lists)?

For 'biologists in general', there has to be some way out of this impasse if genes are to be rescued from the charge of being unreconstructed hooligans flashing their absurd desiderata lists. Have John Maynard Smith and Eörs Szathmáry come to the rescue? After reading *The Origins of Life*, I can't make up my mind. There is an ambivalence in their evolutionary models of the origins of cooperativity that confounds the issue. What is clear is that their starting assumption is pure selfish-genery: "We have to explain how complex entities evolved, despite selection between their components favouring selfish behaviour." Accordingly, this assumption becomes Life's Critical Problem Looking for a Solution, with regards not only to the paradox of the organism, but also to a range of other "major transitions" during evolution. Houston, we have a problem.

There is no point in arguing over the transitions as life on Earth became more 'complex': lonely replicators begat groupie replicators in cells; free-floating replicators begat chromosomes; RNA genes and RNA enzymes begat DNA and proteins; prokaryotes begat eukaryotes; asex begat sex; single cells begat multicells; solitary individuals begat colonies and primate societies begat

human, speaking societies. The interesting arguments concern (i) whether lower-level desires would rip apart higher-level cooperation and (ii) whether there is one overarching principle that gets life through these problematical bottlenecks.

In order to lay out their stall regarding argument (i), Maynard Smith and Szathmáry first need to dismiss 'group selection' which has, over the past two decades, become the *bête noir* of selfish-genists. Group selection does not work, "for individual selection will win out in most cases". However, a recent masterly review of the theory and practice of group selection, *Unto Others: The Evolution and Psychology of Unselfish Behavior* (Harvard University Press) by Elliott Sober and David S. Wilson, shows that it works in ecologically patchy circumstances most likely to be the natural state of affairs for most life-forms. Furthermore, there is an ultimate irony in what Maynard Smith and Szathmáry call their "stochastic corrector model" required under argument (ii), in that it is seemingly rooted in group selection. Indeed, their figure legend of the model ends on the solid note, "group selection can win out", contrary to other statements in the text that "there is no need to invoke group selection" for the transitions.

The stochastic corrector model works as follows: two selfish, free-floating replicators (one type better than the other) find themselves in the first protocells. They begin to replicate, as do the cells, and distribute themselves willy-nilly to new daughter cells. Now comes the trick: insist that only cells inheriting equal numbers of the two replicator types inherit the Earth. Then, plug this little trick into a compliant computer program to see whether what goes in also comes out. The result — "the survival of cooperators" in the establishment of the first cells by group selection.

Group selection versus individual selection (especially when the latter is dumbed-down to genic selection) is not some arcane squabble; it goes to the heart of the phenomenon we call biology, in which thousands of molecules cooperate in bringing to life the all-singing, all-dancing, reproducing organism, passing on the DNA baton from one generation to the next. The ultimate explanation, in my view, will have no more to do with supposedly autonomous selfish replicators than with the vague inner mysteries of holistic biology.

Biology evolved from simple beginnings. We can rid ourselves of the supposedly 'independent' meddlesome replicators if we face up to the reality that there has been, in all



Home is what comes naturally

This is not a fossilized art nouveau dinosaur, but a piece of so-called evolutionary architecture.

The house shown above, created by Eugene Tsui for his parents, is built with materials and with a structure that will withstand earthquakes, flooding, windswept fires and vermin, while a south-facing 'eyeball window' acts as a heat- and

light-magnifying device. But only time will tell whether it is truly adapted to its environment. In his book *Evolutionary Architecture: Nature as a Basis for Design* (Wiley, \$54.95), Tsui suggests that the natural world, for instance a termites' nest or dinosaur bone, can become models for architectural problem-solving.

probability, an intimate, functional 'cross-talk' between consenting molecules of RNA (DNA) and proteins from the Big Origin. Such molecules were small in number, and remain relatively small in number 4.5 billion years later, given the recent findings of rather few (around 2,000) modules (at the DNA and protein levels) shared by approximately 100,000 mosaic genes and proteins. Life is about molecular coevolution between redundant, modular parts, subject to frequent genomic turnover, which have been more than happy to cooperate, in a bewildering variety of combinatorial permutations. Our genes are born to cooperate; they are not full of angry resentment over a 'forced' cohabitation that has been occurring since time immemorial.

If we ditch the selfish-replicator illusion, and accept that the only known biological entity capable of autonomous replication is the cell (full of cooperating genes and proteins, etc.), then we can begin to tackle other events, for example the 'paradox' of the DNA-protein transition stage raised by Maynard Smith and Szathmari. DNA replication is so error-prone that it needs the prior existence of protein enzymes to improve the copying fidelity of a gene-size piece of DNA. "Catch-22," say Maynard Smith and Szathmari. So, wheel on RNA with its now recognized properties of carrying both informational and enzymatic activity, leading the authors to state: "In essence, the first RNA molecules did not need a protein polymerase to replicate them; they replicated themselves." Is this a fact or a hope? I would have thought it relevant to point out for 'biologists in general' that not one self-replicating RNA has emerged to date from quadrillions (10^{24}) of artificially synthesized, random RNA sequences.

So, the "problem" remains; but did it ever exist? Fair play among

mutually interdependent strings of nucleotides and amino acids may have been the only way to produce, together, a self-reproducing entity. This then progressed through all of the transition stages discussed by Maynard Smith and Szathmari and on to the ultimate birth of language — and here we are, arguing the toss over all of this. □

Gabby Dover is in the Department of Genetics, University of Leicester, Leicester LE1 7RH, UK.

An expletive in academia

Merde: Excursions in Scientific, Cultural and Socio-Historical Coprology

by Ralph A. Lewin

Random House: 1999. 157 pp. \$19.95

Malcolm Coe

In 1979, bookshops in Israel purportedly ordered large numbers of copies of Christopher Perrins' serious ornithological work entitled *British Tits*. The title of this book is also eye-catching, but, you might ask, is it a genuine academic subject or an expletive? We are soon in no doubt as to which category it should be in when we are informed that the term 'merde' has its origins in an expression used by the French general Pierre Cambronne in 1815 after his defeat at Waterloo.

The human race has always had a fascination for defecation and death. The two processes have striking ecological similarities, because both result in unwanted material that has been gathered from the surrounding environment and is now on its way back to the ecosystem to be recycled. (It was the great French entomologist Henri Fabre who pointed out the ecological importance

of the recycling roles of arthropods in the mineralization of both dung and corpses.) While much schoolboy humour is strongly faecal (are schoolgirls equally fascinated?), death remains "the one we don't mention".

Defaecation is an inevitable and unavoidable consequence of animal feeding, whether animals possess a true gut or not. And the end-product can tell us a great deal about where the animal has been living, what it has

been feeding on, the quality of its food and, in some cases, how much it has eaten. The hairs of a mammal, the feathers of a bird,

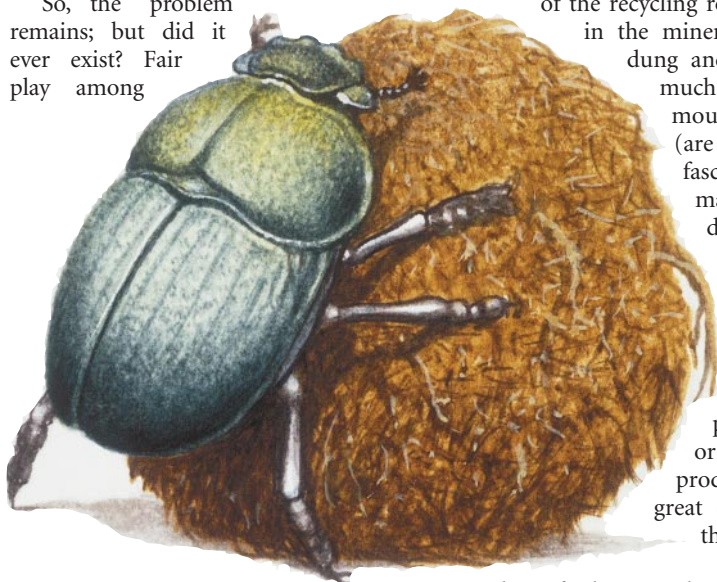
or plant cuticles, all found in faeces, have been used for many years as indicators of feeding habits. Even the fossilized remains of ancient turds known as coprolites may be used to predict diet. Such studies are of considerable interest to forensic scientists, either in relation to dung or to carrion, because the plants, animals and substrates associated with different habitats leave traces that may be powerful indicators of the site of death or deposition.

Humans tend to defaecate in private, probably because crouching during elimination would make them vulnerable to predators or enemies, or transmit disease. But many other mammal species have incorporated this unavoidable activity into their behavioural repertoire. In this case, even squatting during defaecation (or urination) has become ritualized as part of their territorial behavioural signals. Add to this the size, shape, colour and odour of the dung, and we have a powerful source of information on the animals living in a locale. In many cases, of course, animals such as rhinos and hippos may scatter their dung during or after defaecation to enhance its territorial value. Alternatively, dung might be intentionally accumulated in middens to enhance the value of a signal, the site frequently being used by a number of species.

The vigorous activity of dung beetles has impressed people since the time of the Ancient Egyptians, who immortalized the industry of the scarab beetles in their art and religion. In more recent times, Fabre studied the behaviour and biology of a range of dung-beetle genera and species in southern France, demonstrating the numerous niches for different dung types depending on the size and texture of individual droppings. Most of these beetle species, once they have emerged from their pupae, roll or gather dung to feed on until they are sexually mature. They then lay their eggs in dung balls, or 'sausages', that they have buried in their nest chambers. The activity of these insects plays a significant ecological role in environments dominated by large mammals, where their dung-burying activities enhance fertilization, aeration and the percolation of rain into the soil.

Ralph Lewin is clearly a dedicated and entertaining faecophile. It is refreshing to find a scientist who takes neither his subject nor himself too seriously, observing, no doubt to the chagrin of many colleagues, that "It seems a shame that the good name of bullshit, a potentially useful product, should have been debased in recent parlance to signify worthless or misleading statements". Contentiously, the author tells us that the words 'science' and 'shit' are both derived from the same ancient Indo-European root.

The removal of accumulated human waste cannot be ignored in these faecal equa-



Faecophilia: a dung beetle's life begins in dung and is sustained by it.

tions. As human numbers rise ever closer to the point of being unsustainable, it is not surprising that this valuable resource is coming under closer scrutiny. □

Malcolm Coe is at 39 Sandfield Road, Headington, Oxford OX3 7RN, UK.

Dip into the century's technological past

Visions of Technology

by Richard Rhodes

Simon & Schuster: 1999, 400 pp. \$30, £20

Henry Petroski

This anthology opens abruptly, but appropriately, with a view from the late 1920s reminding us of words that were not part of the American vocabulary at the turn of the century — radio, movie, tractor, aviator, income tax, insulin, relativity, quantum theory, behaviourism. The list goes on, and we are engaged immediately by the realization that, throughout the century, most of us have been as close as a generation or two, and at most a few generations away from a very different world of things and ideas.

This collection of excerpts from a broad range of thinking and opinion about technology throughout the twentieth century makes for fast, enjoyable and sometimes fascinating reading. There are first-hand reports of their technologies from the likes of Guglielmo Marconi, Wilbur Wright, Henry Ford and George Eastman; opinions and observations from theorists and visionaries such as Henry Adams, Thorstein Veblen, H. G. Wells and Robert Frost. The authority and stature of these representatives of the earliest years of the century give a fair idea of the quality and breadth of the collection, whose second half, not surprisingly, is filled with discussions of nuclear weapons, the two-cultures debate, the environment and computers. The selections are always apposite, seldom too long, and almost never uninteresting. They average fewer than two pages in length, with many less than half a page.

In addition to textual excerpts, brief quotations, photos, diagrams and graphs are scattered throughout the book, but in general these are not as informative or satisfying as the text. In fact, many of the graphs seem arbitrarily cut off well before the end of the century, frustrating to the reader. One graph, appearing with a 1927 piece by Aldous Huxley on comfort, compares numbers of telephone calls made and pieces of mail sent per capita per year, but is inexplicably curtailed in the mid-1970s. Another graph showing the growth in numbers of personal computers in use from 2 million in 1981 to 38 million in 1987 makes the reader wonder how many are in use today.



A real whodunnit

How many of us, from looking at the two items on the right — a paw track and a snatch of faecal scat — would recognize them as the signature of the bobcat on the left? Paul Rezendes, nature photographer, tracking expert and author of *Tracking and the Art of Seeing: How to Read Animal Tracks and Sign* (HarperCollins, \$24),

would have little difficulty. Bobcats aside, Rezendes' book describes the behaviour and habitats of more than 50 animal species, and how they can be identified, not only from their tracks and droppings, but from trail patterns, scratches and clues that most of us would blunder blithely past.

As with all anthologies, especially wide-ranging ones, any reader will bring his or her own expectations to the book and finish it with a list of addenda and concerns. Engineering as a profession is treated explicitly in fewer than six selections, including one from Herbert Hoover's mid-century memoirs. The editor's introduction to the piece is a bit misleading, however, in stating that "engineering was just becoming a profession at the turn of the century". In spite of the anecdotes of Hoover and others to the contrary, engineering as a profession was, in fact, established well before the beginning of the twentieth century: the American Society of Civil Engineers was founded in 1852 and the British Institution of Civil Engineers some decades before that.

In fact, it is the seemingly ambivalent role of Richard Rhodes as editor of this collection that may be its most distracting and frustrating quality, although the perception may be a result of the book's design. Throughout, the left-hand running head, set in the centre of the left margin and reading vertically over horizontal page numbers, places the name of RICHARD RHODES beside the continuations of quotations from the likes of Edwin Land, Elting Morison, B. F. Skinner and John von Neumann, whose names are not carried

over as running heads.

Some of Rhodes's brief introductions to the selections are helpful, as when he tells us how Gordon E. Moore, one of the founders of the Intel Corporation, came to write the excerpted article that gave us the law relating to the growth rate of integrated circuits. Other selections, however, have such brief introductions that they are almost useless. Thus, in one note to a selection from 1971, Daniel C. Drucker is identified simply as "a dean of engineering". The reader unfamiliar with Drucker's career will not know that he was dean at one of the nation's premier engineering schools, the University of Illinois at Urbana-Champaign. Other authors, generally less well known than Drucker, are left completely unidentified. Perhaps even more disappointing is the fact that the reader is not told the source, occasion or context of the vast majority of the selections. This is an anthology that excels in its selections but that is wanting in its editorial and visual design. □

Henry Petroski is in the Department of Civil and Environmental Engineering, Duke University, Durham, North Carolina 27708-0287, USA. His book on the history of the book shelf, *The Book on the Bookshelf*, will be published by Knopf in September.

A bragger's guide to fuzzy objects

Deep Sky Companions: The Messier Objects

by Stephen James O'Meara
Cambridge University Press: 1999. 318 pp.
£22.50

Stephen P. Maran

As David Levy writes in the foreword to *The Messier Objects*, "There can be no better exercise for a beginning astronomer than to find and observe all of the Messier objects". There are up to 110 of them (depending on how you count them), consisting of star clusters, nebulae and galaxies. They begin with M1, the Crab nebula, and include such other celestial sights as M31, the Andromeda galaxy, and M13, the large globular cluster in Hercules. Most were discovered by Charles Messier (1730–1817), a French astronomer intent on finding new comets.

Given viewing conditions in Paris, it's not surprising that many clusters, nebulae and galaxies would appear as fuzzy blobs, just as a comet may. So Messier conducted the work for which he is remembered, not because he was interested in these things, but to find and tabulate objects that might otherwise confuse his search for comets.

In the past, amateur astronomers trained on Messier objects just as Levy suggests. In the United States, the Astronomical League conducts activities known as the Messier Club and the Binocular Messier Club, and awards certificates for documented personal observations of at least 70 (with telescopes) or 50 (with binoculars) Messier objects (see <http://www.astronleague.org/al/obsclubs/messier/mess.html>).

There are even Messier marathons, held annually in late March or early April (when all the objects are visible), in which contenders try to spot all the Messier objects in a single night. This earns bragging rights but defeats any hope of contemplating the heavens, as no time can be lost in viewing a nebula once it has been identified.

A recent development seems poised to relegate the competitive observation of Messier objects to the same graveyard of discontinued practices as the use of sextants to determine position at sea. What the Global Positioning System does for mariners, the introduction of inexpensive telescopes, equipped with electronic pointing systems and databases of Messier objects and other sky targets is doing for amateur astronomers. They will see sights more readily, but feel less accomplishment in so doing.

For those planning to earn a certificate, "run a marathon" or just enjoy many fine celestial vistas, there is no better guide than Stephen O'Meara's book. For each object, a

photograph, a finder chart, basic data and a quote from Messier's accounts are provided, with a discussion of what is seen through small telescopes, and some history of telescopic studies.

This is not a treatise on the astrophysics of Messier objects, but an observer's guide, a worthy and superior successor to previous attempts. □

Stephen P. Maran is at NASA's Goddard Space Flight Center, Mail Code 600, Greenbelt, Maryland 20771, USA.

Tale of two disciplines

Before Big Science: The Pursuit of Modern Chemistry and Physics, 1800–1940

by Mary Jo Nye
Harvard University Press: 1999. 282 pp.
\$16.95, £10.50 (pbk) [first published 1996]

Bernadette Bensaude-Vincent

There have been numerous complaints over the past decade about the increasing gap between the studies produced by historians of science and the historical accounts that actually interest working scientists. Anyone who has deplored the fact that, while pursuing their own historiographical interests, historians of science have been losing the audience of scientists, will be reassured when they read this volume, the first of a series intended to fill this gap. Here is a concise survey of 140 years of natural science written by a professional historian of science.

Mary Jo Nye clearly targets an educated lay audience with a minimal scientific background. To this end, she has adopted a biographical approach, making extensive use of anecdotes and striking details, including such famous legends as August Kekulé's vision on a London bus of carbon atoms dancing in a circle, which he presented as the origin of the theory of structure.

Clearly, the purpose is to excite the reader's imagination with these heroic figures — but this is by no means a naive, factual historical narrative. Although the title suggests a contemporary perspective, Nye carefully avoids a teleological account of the past that simply describes the formation of contemporary chemistry and physics. On the contrary, her intention is to emphasize the changing profiles of these disciplines and their fluctuating boundaries. Whereas in the mid-nineteenth century, chemistry and physics had developed two independent views of matter, the chemical atoms and molecules being different and disconnected from the physical ones, by the end of the nineteenth century the split was between two chemistries. Organic structural chemistry, which Nye presents as inspired by thinking

from a natural-history standpoint, was opposed to inorganic chemistry, which was based on the Newtonian model of forces. Although spectroscopy and X-ray crystallography confirmed the predictive power of the chemical-structural theories, most organic chemists distrusted physical instruments.

Given this history of mutual disregard, the development of quantum chemistry in the inter-war period appears as a dramatic event. Far from suggesting a necessary reduction of chemistry to quantum physics, Nye concludes this chapter with Linus Pauling's remark: "There is more to chemistry than an understanding of general principles. The chemist is also, perhaps even more, interested in the characteristics of individual substances — that is, of individual molecules."

In stark contrast to the uncertain disciplinary identities, Nye's narrative stresses the strong national identities of scientists. In the late nineteenth century, it was generally admitted that there were national styles of reasoning: the British tradition was said to value pragmatic and visual models, while the German and French traditions liked abstract, more speculative models.

These stereotypes were, of course, influenced by national rivalries. However, they played an active role in this history insofar as they influenced both educational programmes and scientists' self-perception. Despite the huge efforts in the late nineteenth century to set up international meetings and scientific organizations, the scientists portrayed in these pages could not fully define themselves as members of international communities of physicists or chemists. Apart from a few individuals like Albert Einstein, Paul Langevin — the French physicist who invented sonar — and Hermann Staudinger — the Swiss chemist who discovered macromolecules — they remained torn between patriotism and internationalism. In conclusion, Nye stresses the end of the transcendental ethos of the scientist and the more modest self-perception that resulted.

It is thus clear that, far from being merely descriptive, this concise volume is a bold and highly skilled interpretation of the recent past of chemistry and physics. My only regret is that the reasons behind the author's historiographical choices were not explicitly stated in order to convey a better understanding of the methodology used by historians of science. However, this volume provides rich, enjoyable and easy reading. I would recommend it as both an introduction to the history for scientists and a useful tool for those wanting to develop this interest further by using the bibliographical essays in the appendix. □

Bernadette Bensaude-Vincent is in the Department of History and Philosophy of Science, Université de Paris X, 92001 Nanterre, France.

Structure of importin- β bound to the IBB domain of importin- α

Gino Cingolani*, Carlo Petosa*, Karsten Weis† & Christoph W. Müller*

* European Molecular Biology Laboratory (EMBL), Grenoble Outstation c/o ILL BP 156, 38042 Grenoble Cedex 9, France

† University of California, Berkeley, Department of Molecular & Cell Biology, Berkeley, California 94720-3200, USA

Cytosolic proteins bearing a classical nuclear localization signal enter the nucleus bound to a heterodimer of importin- α and importin- β (also called karyopherin- α and - β). The formation of this heterodimer involves the importin- β -binding (IBB) domain of importin- α , a highly basic amino-terminal region of roughly 40 amino-acid residues. Here we report the crystal structure of human importin- β bound to the IBB domain of importin- α , determined at 2.5 Å and 2.3 Å resolution in two crystal forms. Importin- β consists of 19 tandemly repeated HEAT motifs and wraps intimately around the IBB domain. The association involves two separate regions of importin- β , recognizing structurally distinct parts of the IBB domain: an amino-terminal extended moiety and a carboxy-terminal helix. The structure indicates that significant conformational changes occur when importin- β binds or releases the IBB domain and suggests how dissociation of the importin- α / β heterodimer may be achieved upon nuclear entry.

Macromolecular traffic between the cytoplasm and the nucleus is a fundamental activity of eukaryotic cells^{1–3}. Exchange between the two cellular compartments occurs through the nuclear pore complex (NPC), a structure of relative molecular mass 125,000,000 ($M_r = 125,000K$), containing around 100 different polypeptides, and embedded in the nuclear envelope^{4,5}. Targeting of proteins to the nucleus is determined by the presence of a nuclear-localization sequence (NLS), typically composed of one or more clusters of basic amino-acid residues⁶. The single cluster in the SV40 large T antigen (¹²⁶PKKKRKV¹³²)⁷ and the two in nucleoplasmin (¹⁵⁰KRPAAIKKAGQAKKK¹⁷⁰)⁸ are well-characterized examples of a monopartite and a bipartite NLS, respectively. An NLS-bearing substrate is delivered to the nucleus by association with the heterodimer formed by importin- α (also called karyopherin- α)^{9–11} and importin- β (karyopherin- β)^{12–15}. Importin- α recognizes the NLS, and importin- β is responsible for docking to the NPC and translocation through the pore. Dissociation of the NLS–importin- α / β trimer is triggered by the association of importin- β with Ran, a Ras-related GTPase. Ran is predominantly bound to GTP in the nucleus and to GDP in the cytoplasm. This uneven distribution leads to the correct spatial coordination of NLS-substrate release, because importin- β is recognized by Ran-GTP but not by Ran-GDP¹⁶. The interaction of importin- β with Ran-GTP is tight ($K_D = 0.5$ nM) and inhibits GTP hydrolysis¹⁷.

Importin- α ($M_r = 60K$) is composed of a small N-terminal domain and a large NLS-binding domain connected by a flexible linker. The crystal structure of the NLS-binding domain from yeast importin- α comprises ten armadillo (arm) repeats, with each arm being composed of three helices¹⁸. The arm repeats are stacked to form an elongated right-handed superhelix containing a shallow groove where the NLS peptide is bound. The N-terminal domain, termed the importin- β -binding (IBB) domain, is a basic stretch of about 40 highly conserved residues, which are minimally required for binding importin- β and for efficient nuclear entry^{19,20}. The IBB domain is thus the nuclear targeting signal of importin- α and has been suggested to be evolutionarily related to classical NLS signals¹. In the crystal structure of murine importin- α , the IBB domain is mostly disordered. However, a short stretch of basic residues is bound to the NLS-binding site, suggesting a role for the IBB domain in the autoinhibition of importin- α ²¹.

Importin- β ($M_r = 97K$) has Ran-, IBB- and NPC-binding activities. The interactions with Ran and the IBB domain have been mapped to N-terminal and C-terminal regions, respectively^{22–24}. In contrast, the residues involved in NPC binding are unknown.

Importin- β is the best characterized member of a superfamily of homologous proteins involved in nuclear transport, which include CAS, transportin and exportin-1 (ref. 5). The relative molecular masses of members of this superfamily vary between 90K and 130K and are characterized by an N-terminal Ran-binding region and by NPC-binding activity²⁵. Importin- β contains a number of tandem HEAT motifs, sequence repeats of roughly 40 residues found in a variety of eukaryotic proteins²⁶. In the crystal structure of one of these, the PR65/A subunit of protein phosphatase 2A, fifteen tandemly repeated HEAT motifs are assembled into an elongated, left-handed superhelix²⁷. The individual HEAT motif forms a helical hairpin structure reminiscent of an armadillo repeat, consistent with the suggestion that arm and HEAT-repeat-containing proteins are evolutionarily related.

Apart from the classical protein-import pathway, importin- β is also involved in the nuclear import of spliceosomal small nuclear ribonucleoprotein particles (snRNPs)²⁸, which is mediated by the 45K protein snurportin-1 (ref. 29). Snurportin-1 is functionally analogous to importin- α , as it recognizes the trimethylguanosine (m_3G) cap structure of the small nuclear RNAs U1, U2, U4 and U5, and binds to importin- β through an IBB domain. Furthermore, recent studies show that a number of proteins are imported into the nucleus bound directly to importin- β , without binding to importin- α . These include ribosomal proteins³⁰, CDK/cyclin complexes³¹ and the viral proteins HIV-1 Rev, HIV-1 Tat³² and HTLV-1 Rex³³.

We have solved the crystal structure of human importin- β bound to the IBB domain of human importin- α . The association occurs through a combination of specific-sequence recognition and gross electrostatic complementarity. The structure provides clues as to how the binding of Ran-GTP to importin- β could lead to IBB dissociation, and supports the hypothesis that importins α and β have evolved from a common ancestor³⁴.

Structure determination

Extensive attempts to crystallize importin- β either by itself or in complex with importin- α were unsuccessful. In proteolytic assays, strong proteolytic protection of importin- β in the presence of importin- α indicated an altered, possibly more compact conformation of the molecule (data not shown). Similar protection was observed when importin- α was replaced by a chemically synthesized peptide spanning the IBB domain (residues $\alpha 11$ – $\alpha 54$). (We denote residues in importin- α and in the IBB domain with the prefix α to distinguish them from residues in importin- β .) Co-crystallization experiments of importin- β with the IBB domain gave rise to two

Table 1 Structure determination of the Importin- β -IBB domain complex

Native data statistics											
Data set	Space group	a (Å)	b (Å)	c(Å)	β (°)	Resolution (Å)	Completeness (%)	Multiplicity (%)	<i>R</i> _{sym} * (%)	<i>I</i> / <i>σ</i>	ESRF beam line
Crystal form I	P2 ₁	64.9	97.6	83.2	91.2	30-2.5	97.5 (97.5)	2.5	7.1 (30.2)	15.7 (2.5)	ID02
Crystal form II	P2 ₁	57.9	101.2	83.8	92.1	40-2.3	98.5 (98.5)	7.7	8.9 (38.0)	13.7 (2.0)	ID14-3
Statistics of the MAD data (Crystal Form I)											
MAD	Resolution (Å)	Completeness (%)		<i>R</i> _{sym} * (%)		Number sites found/total		Figure of merit	Z-score†	ESRF beam line	
Inflection point λ1 (0.9791 Å)	30-3.3	97.4 (98.6)		7.5 (25.3)		22/24		0.59	99.9	BM14	
Absorption peak λ2 (0.9790 Å)	30-3.3	97.1 (98.1)		8.9 (27.6)							
Remote point λ3 (0.8856 Å)	30-3.3	97.2 (97.8)		8.5 (25.0)							
Refinement statistics											
					Crystal form I		Crystal form II				
Resolution range (Å)					30-2.5		40-2.3				
Total number of non-hydrogen atoms					7185		7011				
Number of water molecules					44		186				
<i>R</i> -factor (%)					22.1 (33,029 reflections)		23.5 (43,760 reflections)				
<i>R</i> _{free} (%)‡					27.3 (1,759 reflections)		28.5 (2,535 reflections)				
r.m.s. deviation from ideal bond length (Å)					0.007		0.007				
r.m.s. deviation from ideal bond angles (°)					1.2		1.2				

Numbers in bracket are reported outer shell statistics.
 $R_{\text{sym}} = \sum_i \sum_h |I(i, h) - \langle I(h) \rangle| / \sum_i \sum_h I(i, h)$ where $I(i, h)$ and $\langle I(h) \rangle$ are the i th and mean intensity of reflection h .
 $\dagger$ Z-score, quality criteria according to program SOLVE⁴⁶.
 $\dagger$ R_{free} is calculated using 5% of the reflections.

related crystal forms in space group $P2_1$, both diffracting beyond 2.5 Å resolution. The structure of the complex was solved in crystal form I by the multiwavelength anomalous diffraction (MAD) method using selenomethionine-substituted importin- β , and was refined at 2.5 Å resolution (Table 1). A section of the final electron-density map is shown in Fig. 1d. Crystal form II was solved by molecular replacement and the model was refined to 2.3 Å resolution. In crystal form I, the model includes all residues ($\alpha 11$ – $\alpha 54$) of the IBB domain and all 876 residues of importin- β . In crystal form II, it includes IBB residues $\alpha 27$ – $\alpha 54$ and all importin- β residues except for residues 615–620, which form a disordered loop. Our discussion of the IBB-domain structure and its recognition by importin- β is based on crystal form I.

Overview of the complex

The structure of the complex is shown in Fig. 1. The complex takes on a snail-shaped appearance with the IBB domain at the centre and importin- β wrapped around the outside. The complex adopts a highly compact, essentially globular shape with a diameter of 85 Å. Importin- β is an all-helical protein composed of 19 tandem HEAT repeats, arranged in a right-handed superhelix with both a pitch and a mean diameter of approximately 50 Å. The relatively short pitch results in N- and C-terminal HEAT repeats nearly touching each other, giving the molecular a rather ‘closed’ appearance. The shape of the molecule is thus dramatically different from that of the phosphatase 2A PR65/A subunit, which forms a left-handed superhelix and has a much more open conformation (Fig. 1c). HEAT repeats 7–19 are responsible for the recognition of the IBB domain, while HEAT repeats 1–6 are not involved in the IBB binding interaction but are implicated in the interaction with Ran.

The IBB domain is intimately bound on the inner surface of importin- β . The domain has two structurally distinct parts: an N-terminal extended moiety (residues $\alpha 11$ – $\alpha 23$) and a C-terminal helix ($\alpha 24$ – $\alpha 51$), whose axis is roughly coincident with that of the importin- β superhelix. The N-terminal moiety interacts with helices in HEAT repeats 7–11 and with an acidic loop in HEAT repeat 8 (HEAT-8), while the IBB helix packs against helices from HEAT repeats 12–19. The inner surface of importin- β contains many acidic residues and is thus complementary to the highly

positively charged IBB domain.

Structure of importin- β

Each HEAT repeat in importin- β is composed of an A and a B helix connected by a short turn of residues. A representative HEAT repeat, HEAT-10, is shown in Fig. 2c. Repeats are arranged within the molecule to produce an outer layer of A helices defining the convex surface and an inner layer of B helices defining the concave surface. Neighbouring repeats are joined by residues forming a right-handed connection, the ‘linker’. The HEAT repeats vary between 32 and 61 residues in size, reflecting variations in helix A (3–6 helical turns), helix B (3.5–8 turns) and the linker (1–19 residues). The two helices have a mean crossover angle of 32°, typically with helix B longer than helix A by at least one helical turn. The turn joining the helices is 3–6 residues long in most repeats, and the sharp change in chain direction is often facilitated by a proline at position 19. (The numbering convention used is that of the PR65/A consensus HEAT motif, as shown in Fig. 2b.) In a few repeats, this turn is replaced by a hydrophilic loop, the most conspicuous example being a highly acidic loop (³³⁷DDDDDW³⁴²) in HEAT-8, with five consecutive aspartate residues followed by a strictly conserved tryptophan involved in IBB-domain recognition. The inter-repeat linker is in a random-coil conformation, occasionally with a single turn of 3_{10} helix. The only exception is the linker between repeats 2 and 3, which contains an α -helix 12 residues long.

The nineteen HEAT repeats share a common core of 21 residues, comprising about 3 turns of A helix and 4 turns of B helix. These can be pairwise superimposed with an r.m.s. deviation (r.m.s.d.) of 0.7–1.3 Å for C α atoms. Despite this structural homology, the HEAT repeats, share very low sequence identity (the average for all pairwise comparisons is 12% identity). Nevertheless, the alignment in Fig. 2b shows that the chemical nature of residues is generally conserved. In particular, hydrophobic residues are preferred at positions 9, 10, 13 and 17 in helix A and at positions 24, 28, 31, 32, 35 and 36 in helix B, in agreement with the PR65/A consensus sequence. These residues are involved in packing interactions within and between repeats. Consensus residues Leu(13) in helix A and Ala(28) in helix B are particularly well conserved (boxed in Fig. 2b, c). Position 15 on helix A is a highly conserved polar or charged

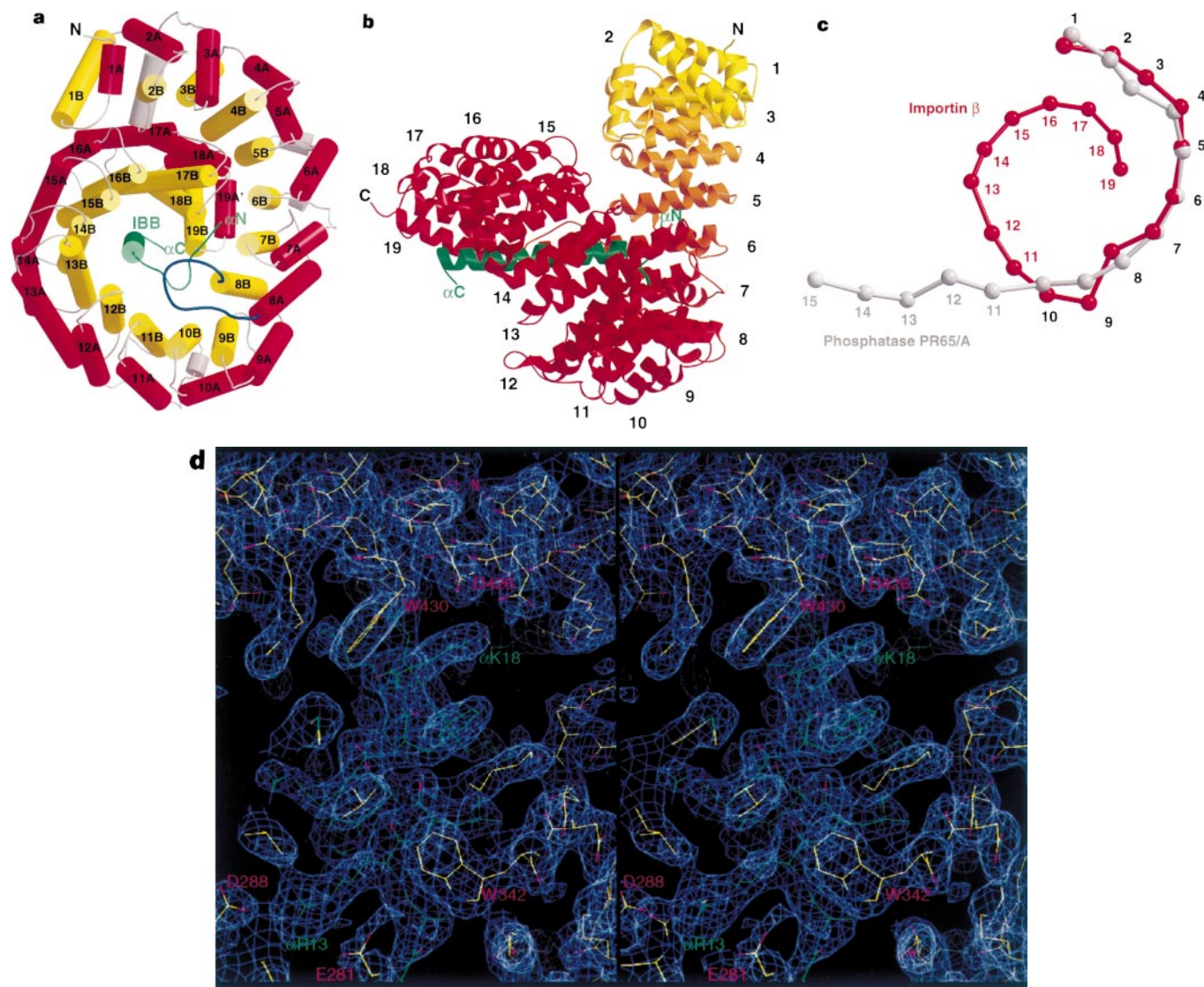


Figure 1 Structure of importin- β bound to the IBB domain of importin- α . **a**, View down the superhelical axis. A and B helices are shown in red and yellow, respectively, with connecting residues in grey. The acidic loop with the DDDDDW motif in HR8 is shown in blue. **b**, Side view. The ribbon representing importin- β is coloured progressively from yellow to red as the chain proceeds from N to C terminus. The model has been rotated by 90° about a vertical axis with respect to

a, c, Comparison of importin- β and the protein phosphatase 2A PR65/A subunit. Proteins are represented as 'beads-on-a-string', with the centre of each HEAT repeat indicated by a sphere. **d**, Final $2F_o - F_c$ electron-density map (contoured at 1.0 σ) of the N-terminal IBB moiety in crystal form I. Carbon atoms of the IBB domain and importin- β are in green and yellow, respectively. Figs 1a–c, 4 and 5 were prepared with MOLSCRIPT⁴⁸.

residue, most often a glutamate. In almost all the HEAT repeats, this position is exposed to solvent and is not involved in packing interactions. A relatively conserved basic residue at position 25 is partially solvent accessible at the top of helix B. However, the extended network of interactions among highly conserved Asp(19) and Arg(25) residues observed in the PR65/A subunit is absent from importin- β .

The A helices in HEAT repeats 7–10, 12–16 and 18 are significantly bent. The curvature is always in the same direction, with the ends of the helix bending towards the interior of the molecule. HEAT-19 is unusual in that the A helix is replaced by two short helices (19A and 19A') oriented at 90° to each other. In HEAT repeats 8–10 and 15, a proline at position 11 accounts for a ~40° kink in the A helices. This corresponds to the highly conserved proline in the HEAT motifs of the PR65/A subunit²⁷. A proline performing a similar role is found offset towards the N-terminus by one or two helical turns in HEAT repeats 7, 16 and 18, and between helices A and A' in HR19.

Neighbouring HEAT repeats are stacked ~11 Å apart and are

related by a rotation and by a tilt, which together generate the curvature of the molecule. The tilt equates to adjacent A helices being spaced further apart than adjacent B helices, and arises primarily because A helices use large hydrophobic residues to pack against adjacent helices, whereas B helices prefer smaller residues like valine or alanine, reminiscent of the effect observed in ankyrin-repeat-containing proteins. The rotation occurs about an axis approximately parallel to a line joining the centres of mass of adjacent HEAT repeats. In most cases this rotation is 10° to 25° clockwise as one proceeds from the N to the C terminus, although it is nearly 0° between HEAT-6 and 7 and between HEAT-12 and 13. In three cases, the rotation is 30° to 50° counterclockwise: between HEAT-1 and 2, HEAT-3 and 4, and HEAT-8 and 9. THE HEAT-1 repeat is splayed away from the rest of the molecule, with helix 1A packed orthogonally against helix 2B, and helix 1B largely exposed to the solvent. The packing between HEAT-8 and 9 is noteworthy because it breaks a pattern that extends from repeats 4 to 19. This can be seen in Fig. 1a: as one proceeds from HEAT-4 to 8, the B helices are tilted increasingly toward the centre, but helix 9B breaks

The IRB domain is an L-shaped molecule which spans ~ 45 Å from

The IBB domain is an L-shaped molecule which spans 115 Å from end to end, with the N-terminal moiety ($\alpha 11$ – $\alpha 23$) and the C-terminal helix ($\alpha 24$ – $\alpha 51$) running in mutually perpendicular directions. Residues $\alpha 52$ – $\alpha 54$ are in an extended conformation. All the residues of the IBB domain are well ordered, with the average atomic *B* factor ($59 \text{ \AA}^2$) only slightly higher than that of the protein ($52 \text{ \AA}^2$). Basic residues account for nearly 40% of the IBB domain, creating an electrostatic surface potential complementary to that of

HEAT-6 (Fig. 3b, c). Most residues in the IBB helix are in contact with residues on the B helices of HEAT repeats 12–19 (Fig. 3b, d). Notably, the six acidic residues in the IBB domain are all located on one side of the IBB helix and, apart from α Asp 44, do not interact with importin- β .

Residues α Arg 13 and α Lys 18 are the only residues in the IBB N-terminal moiety that are strictly conserved across importin- α homologues (Fig. 2a). Their recognition occurs through a spatial

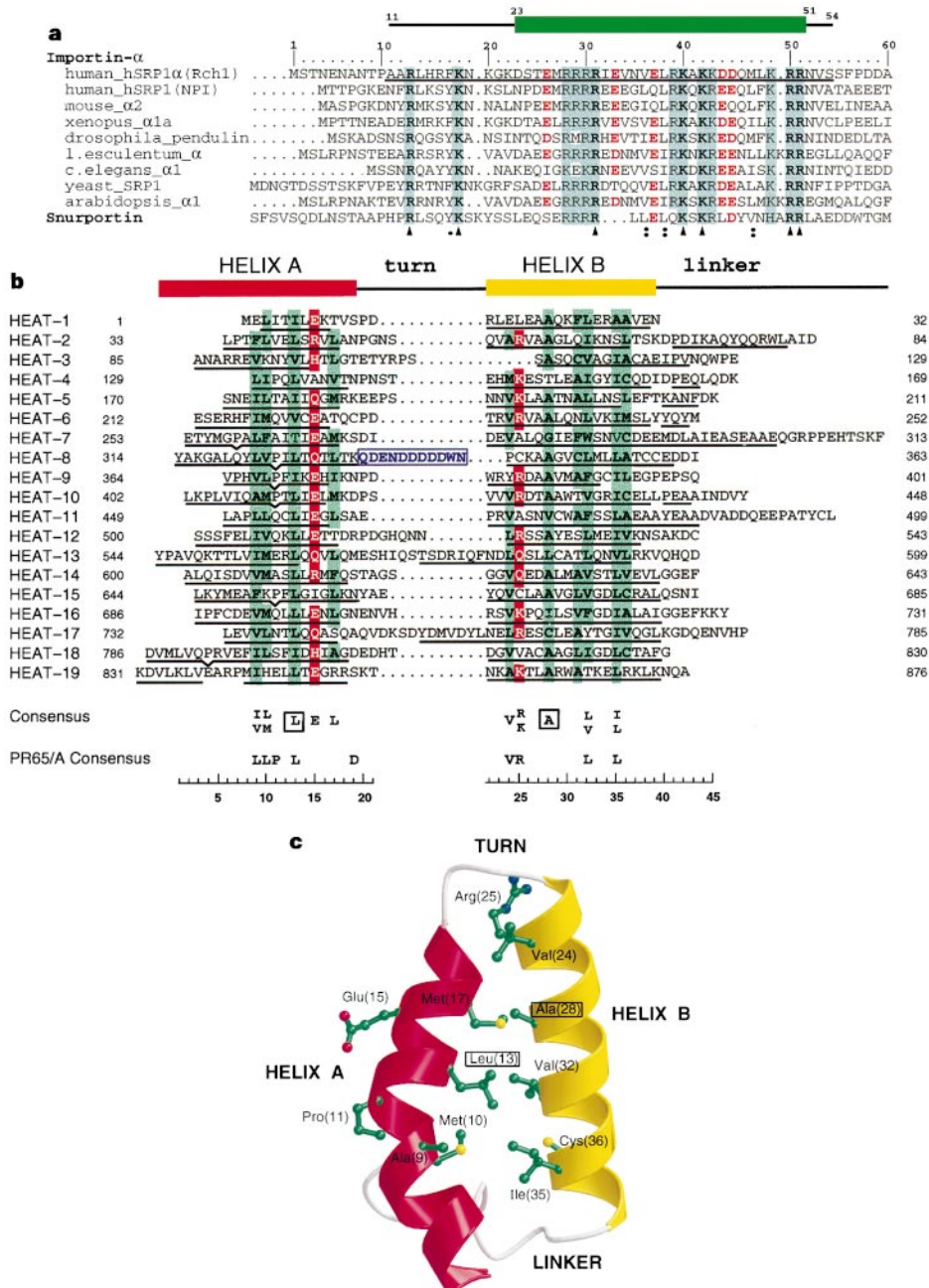


Figure 2 Primary and secondary structures of importin- β and the IBB domain. **a**, The IBB domain of human importin $\alpha 2$ aligned with eight homologues, including human snurportin1. Conserved residues are in red (acidic) or highlighted in blue (basic). Invariant residues are indicated by a black triangle, highly conserved hydrophobic residues by a colon, and α Phe(Tyr)17 by a dot. **b**, Structural alignment of the 19 HEAT repeats. Conserved hydrophobic and polar residues are

highlighted in green and red, respectively. The acidic loop is enclosed in a blue box. A line underneath residues indicates a helix, and a kink in the line marks a kink in the helix. **c**, A representative HEAT repeat. Consensus residue positions in HEAT-10 are shown. Leu(13) is actually Leu413. Highly conserved residues Leu(13) and Ala(28) are boxed as in **b**.

arrangement of highly conserved acidic and hydrophobic residues. The side-chain methylene groups of α Arg 13 pack against the side chains of Val 350 on helix 8B and the strictly conserved Trp 342 of the acidic loop. Trp 342 also interacts with main-chain and side-chain atoms of residues α 12– α 14, which run parallel to the plane of the indole group. The guanidinium group of α Arg 13 forms a bidentate salt bridge with the strictly conserved Glu 281 and the highly conserved Asp 288 (always an aspartate or a glutamate). This may explain the preference for arginine at position α 13, as a shorter lysine side chain would interact with only one acidic residue.

Residue α Lys 18 is in an unusual backbone conformation caused by a sharp bend in chain direction, with ϕ - and ψ -angle values of 48° and -118° , just bordering a generously allowed region in the Ramachandran plot. Many interactions stabilize this residue: the amino group forms a salt bridge with the highly conserved Asp 426 and a hydrogen bond with the O_γ atom of Thr 427; the side-chain methylene groups are in Van der Waals contact with Met 388 and are sandwiched between the aromatic rings of the invariant residue Trp 430 and the highly conserved α Phe 17 (always a phenylalanine or a tyrosine). Additional interactions involving the N-terminal IBB

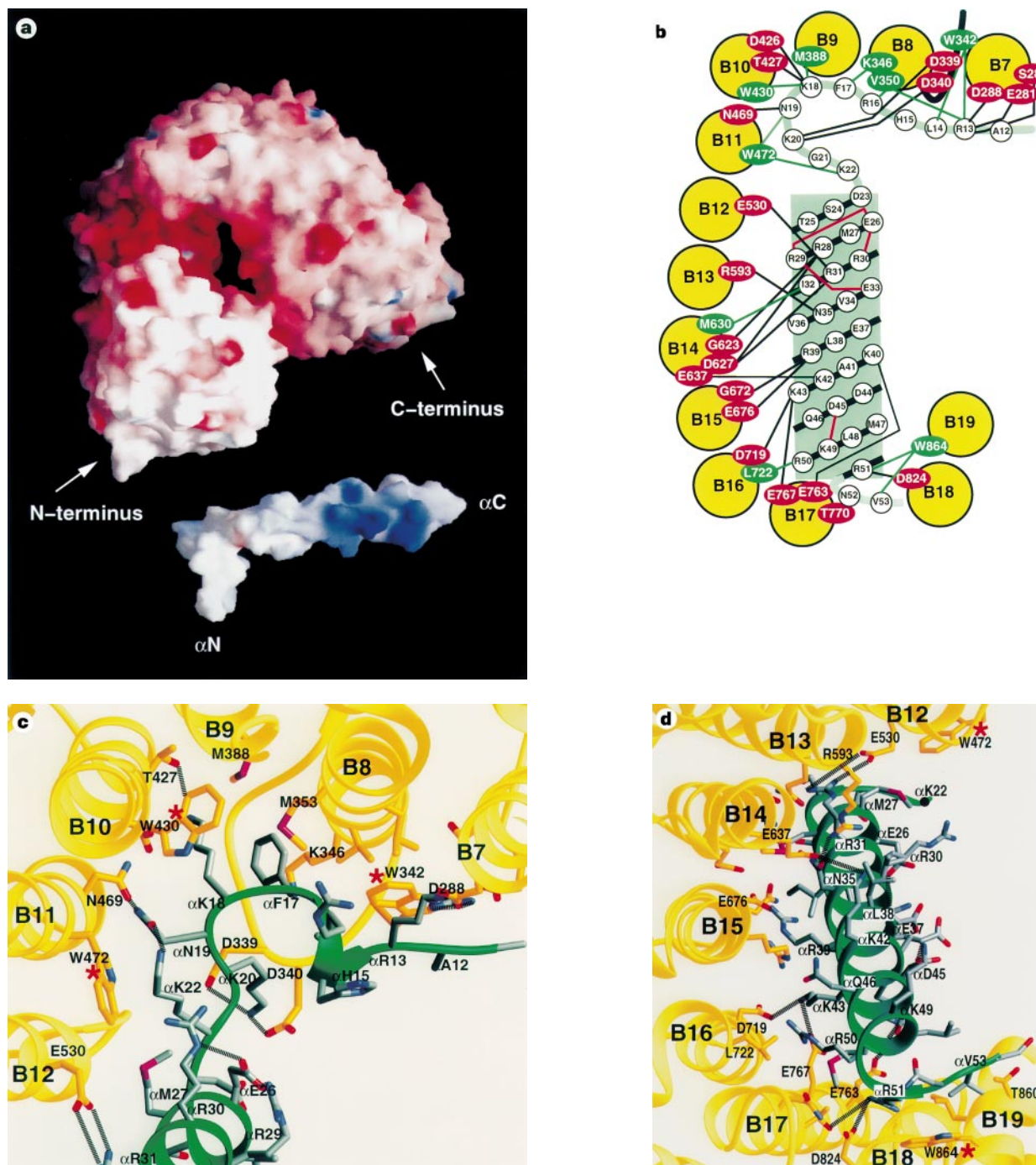


Figure 3 Recognition of the IBB domain by importin- β . **a**, Electrostatic surface representation of the importin β -IBB domain complex. Blue and red regions represent positive and negative potential. Importin- β wraps around the IBB-domain, which has been translated vertically to the bottom of the figure. **b**, A summary of interactions. Importin- β residues making polar and hydrophobic interactions are shown in red and green, respectively. Intermolecular polar and

hydrophobic interactions are indicated by black and green lines. Intramolecular salt bridges of the IBB domain are depicted in red. **c**, **d**, Interactions involving the N-terminal moiety (**c**) and the C-terminal helix (**d**) of the IBB domain. Polar interactions are indicated by dashed lines. Tryptophan residues W342, W430, W472 and W864 are marked with red asterisks. Figures were produced with programs GRASP⁴⁹ and RIBBONS⁵⁰.

moiety include a hydrogen bond between the side chains of α Asn 19 and Asn 469; a salt bridge between α Lys 20 and residues Asp 339 and Asp 340 in the acidic loop; and Van der Waals interactions between α Lys 22 and Trp 472. α His 15 does not contact any importin- β residues, consistent with its poorly conserved nature.

The B helices of HEAT repeats 12–19 wrap 270° around the C-terminal helix of the IBB domain (Figs 1, 3d) and are inclined by ~30° with respect to the IBB helix. The IBB helix is sufficient to mediate complex formation, as a fragment spanning residues α 21– α 65 which lacks the N-terminal moiety can still bind importin- β , although with much reduced affinity²⁰. Of the 16 charged residues in the IBB helix, 13 are involved in salt bridges. Most of these involve highly conserved aspartate or glutamate residues of importin- β and are either monodentate (α R28 with D627, α R39 with D676, α K42 with E637, α R51 with D824) or bidentate (α R31 with E530 and D627, α K43 with D719 and E767). There is also an example in which the charge pair is reversed (α D44 with K857 on helix B19). The recognition of α Arg 51 is like that of α Arg 13 and α Lys 18 in the N-terminal moiety. The guanidinium group forms a salt bridge with the strictly conserved Asp 824 whereas the side chain is boxed by van der Waals interactions, sandwiched between the strictly conserved Trp 864 and α Met 47, and between the C α atom of Gly 820 and side chains of Thr 770 and Val 861. Residue 820 is a conserved glycine or alanine residue, as a larger side chain would not accommodate α Arg 51.

The involvement of HEAT repeats 7–19 in IBB domain recognition and the exclusion of repeats 1–6 from the interaction agree well with experiments in which importin- β deletion mutants were tested for importin- α -binding affinity: a mutant spanning residues 256–876 (repeats 7–19) binds with wild-type affinity, one spanning residues 331–876 (repeats 8–19) binds with much reduced affinity, and one missing residues 772–876 (repeats 18 and 19) fails to bind²². The highly specific interactions involving α Arg 13 and α Lys 18 agree well with mutagenesis data: deletion of residues up to but not including α Arg 13 yields an importin- α mutant that displays wild-type activity in an *in vivo* import assay, whereas deletion of the additional arginine residue reduces activity by 50%; deletion up to and including α Lys 18 further decreases activity three-fold¹⁹. The interactions of α Arg 13, α Lys 18 and α Arg 51 with highly conserved acidic and tryptophan residues are strikingly

reminiscent of those mediating the recognition by importin- α of basic residues in an NLS ligand¹⁸ (discussed below).

Implications for conformational change

The closed conformation of importin- β and the eight-HEAT-repeat 'embrace' of the IBB helix suggests that a marked conformational change may occur when the IBB domain is either bound or released. This is consistent with preliminary data indicating that importin- β is stabilized in a protease-resistant state upon association with the IBB domain (not shown), and agrees with our inability to crystallize importin- β in the absence of the IBB domain, even though the latter is not involved in any crystal contacts. We suspect that importin- β adopts a more open conformation when not bound to the IBB domain, perhaps similar to that of the PR65/A subunit of phosphatase 2A (Fig. 1c). Our second crystal form provides some insight into the flexibility of importin- β (see below).

Crystallographic and biochemical data indicate that the IBB domain also changes conformation upon complex formation. In the structure of mouse importin- α , residues α 44– α 54 lie within the NLS-binding site in an extended conformation²¹. However, in our structure these residues form two turns of helix plus three residues in an extended conformation at the C-terminus of the IBB domain. Also, chymotrypsin cleaves importin- α at residues α Arg 28 and α Asn 35 in the absence of importin- β , suggesting that these residues are poorly ordered; however, in our structure they occur within the IBB helix. Thus, the IBB helix is probably several helical turns shorter when not bound to importin- β . The IBB domain has a stretch of highly conserved arginines, α 28RRRR α 31, of which only α Arg 28 and α Arg 31 interact with importin- β . α Arg 29 and α Arg 30 are instead involved in intramolecular salt bridges with α Glu 26 and α Glu 33, at the base of the IBB helix (Fig. 3b, d). Another salt bridge is observed at the opposite end of the helix, between α Asp 45 and α Lys 49. These salt bridges link residues in successive turns of the IBB helix, and might be important for stabilizing the helix when bound to importin- β .

Comparison of the two crystal forms

The large number of specific intermolecular contacts and the ability to account for sequence conservation and mutagenesis data makes us confident that the conformation observed in crystal form I is

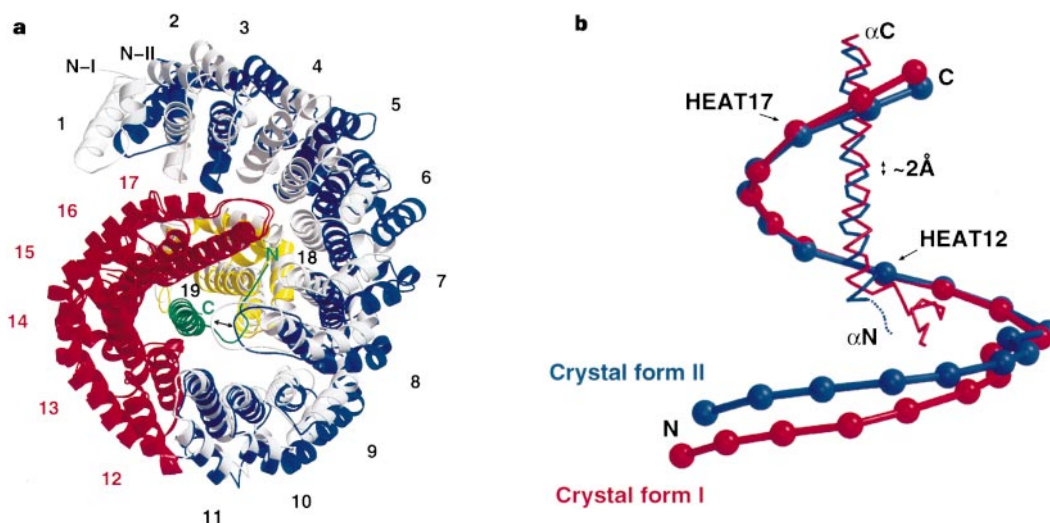


Figure 4 Comparison of the complex in two crystal forms. Three regions of the molecule can be reasonably superimposed: HEAT repeats 1–11 (r.m.s.d. for 494 C α atoms = 0.7 Å); HEAT repeats 12–17 (r.m.s.d. for 246 C α atoms = 1.0 Å) and HEAT repeats 18–19 (r.m.s.d. for 117 C α atoms = 0.6 Å). **a**, View along the superhelical axis. The central rigid body region used for the superposition is shown in red. For crystal form II the N- and C-terminal regions are shown in blue

and yellow, respectively; for crystal form I they are in grey. Only the IBB domain of crystal form I is shown. The shift in the acidic loop of HR18 is indicated. **b**, Beads-on-a-string representation. The more compact nature of crystal form II is clear. The shift in the IBB helix is indicated. Disorder in the N-terminal IBB moiety in crystal form II is indicated by a dashed line.

biologically relevant. In contrast, the conformation in crystal form II may be partly due to experimental conditions. Crystals of this form were grown at low pH (pH 3.9) and were deliberately allowed to become partially dehydrated, a technique leading occasionally to better diffraction quality³⁵; in fact, crystal form II diffracts better than form I, to 2.3 Å rather than to 2.5 Å. Both crystal forms belong to the same space group and have relatively similar unit cell parameters (Table 1). The molecular packing arrangement is also approximately the same, although tighter in crystal form II, consistent with a lower percent solvent content (48.1% versus 51.4%).

The overall shape of importin-β is similar in the two crystal forms, but there are substantial differences in conformation (the r.m.s.d. for all Cα atoms is 2.9 Å). The conformational change, as analysed by program DYNDOM³⁶, consists essentially of two rigid body movements, a ~10° rotation of HEAT repeats 1–11 with respect to the central HEAT repeats 12–17, and a similar rotation of repeats 18 and 19, such that atoms in HEAT repeats 1 and 19 move by as much as 10 Å and 5 Å, respectively (Fig. 4). These movements result in a superhelix with a shorter pitch and a larger radius of curvature. The molecule has a higher degree of intramolecular contact between N- and C-terminal HEAT repeats and is, on the whole, more compact (the gyration radius is smaller by 1.4%). The movement of repeats 1–11 affects residues involved in binding the IBB N-terminal moiety. In particular, helix B and the acidic loop in HEAT-8 move by almost 5 Å, so that they are further away from the HEAT repeats involved in recognizing the IBB helix (Fig. 4a). Trp 472, which in crystal form I interacts with αLys 22, adopts a different rotamer conformation in crystal form II, placing the indole ring far from the αLys 22 binding site.

Changes also occur in the IBB domain. In crystal form II, N-terminal residues α11–α26 are poorly ordered and have not been modelled. Residues in the IBB helix are ordered but their B-factors are unusually high (75–90 Å²). The disorder is probably due to the altered conformation of importin-β, or perhaps a weakening of salt bridges at the low pH employed. The IBB helix is shifted along its axis by 2 Å relative to HEAT repeats 12–17 (Fig. 4b). The helix moves in unison with HEAT repeats 18 and 19, preserving the local environment of the conserved αArg 51 residue. Most salt-bridge interactions between the IBB helix and HEAT repeats 12–17 are maintained. However, many interatomic distances are altered and a few side chains change conformation, including that of αArg 31, located between Glu 530 and Glu 627 in crystal form I, but between Glu 626 and Glu 627 in crystal form II. Thus, some variation in the position of the IBB helix and the conformation of its side chains appears to be tolerated.

The conformation in crystal form II permits several observations.

First, HEAT repeats can move relative to one another, showing that importin-β has a degree of flexibility. Second, the two moieties of the IBB domain appear to interact with their cognate HEAT repeats independently, indicating that the two moieties might be bound or released in a sequential fashion. Third, rigid body movements can lead to an altered spatial separation of the two HEAT-repeat regions which recognize the two IBB moieties. A conformational change in importin-β (possibly induced by binding to Ran or other factors) could facilitate dissociation of the IBB domain by making it impossible to maintain interactions simultaneously with both IBB moieties. Furthermore, the acidic loop in HEAT-8 is stabilized through interactions with residues in HEAT repeats 9–12, and hence may be particularly sensitive to changes in overall conformation. Because this loop contains the conserved Trp 342 residue involved in recognizing the N-terminal IBB moiety, a conformational change of the loop could lead to a weaker IBB binding interaction.

Interactions with the NPC and with Ran

Biochemical data indicate that HEAT repeats 15–19 are not required for NPC-binding activity, as a fragment comprising residues 1–618 binds the NPC with wild-type activity²². Also, at least two distinct NPC-binding sites appear to exist. One site may be located in HEAT-7, as the C-terminal fragment spanning HEAT repeats 7–19 binds the NPC more strongly than one spanning HEAT repeats 8–19, and the N-terminal fragment containing HEAT repeats 1–7 can still bind the NPC while one containing HEAT repeats 1–6 cannot²². We examined the spatial distribution of highly conserved residues to check whether they indicated a possible NPC-binding site on the protein surface. Fifteen per cent of the 876 importin-β residues are strictly conserved across species, including yeast, *Drosophila* and humans. Almost all are buried in the hydrophobic core or are involved in IBB-domain recognition. A few are exposed on the protein surface, but these are scattered and do not form a conspicuous patch. Thus, it remains unclear what features of importin-β are recognized by the NPC.

In contrast, Ran-binding activity has been shown to reside within HEAT repeats 1–8 (refs 22, 24). Binding activity is greatly reduced upon deletion of the first 7, 10 or 33 residues in HR1 (refs 22, 24), and is eliminated upon further deletion to residue 44 at the end of helix 2A²⁷. HR1 is prominently splayed away from the rest of the molecule and is probably an important determinant of Ran binding. Whereas a fragment containing residues 1–364 can bind Ran, one containing residues 1–343 cannot²². The additional deletion removes helix 8B and compromises the integrity of the acidic loop containing the DDDDDW motif. Although the acidic nature

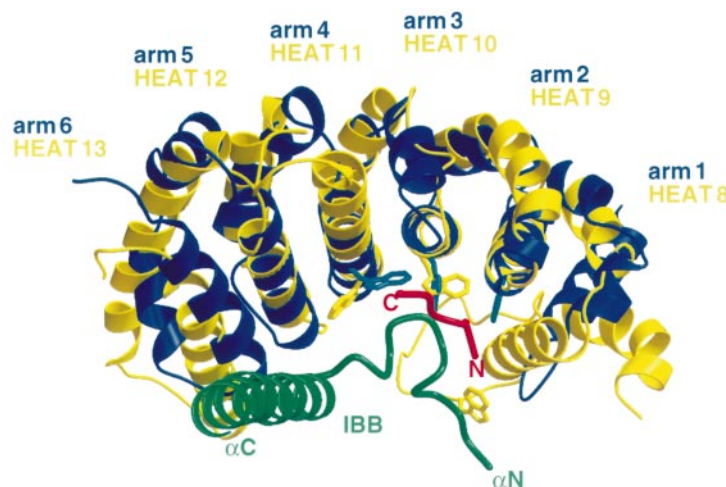


Figure 5 A comparison of importins α and β. Armadillo repeats 1–6 of importin-α (blue) are superimposed on HEAT repeats 8–13 of importin-β (yellow). The NLS

peptide and IBB domain are depicted in green and red, respectively. Tryptophans involved in substrate binding are shown.

of this loop is highly conserved, the aspartates do not engage residues in the IBB domain, except for the poorly conserved residue α Lys 20. Moreover, the acidic stretch is present in the sequence of other Ran-binding homologues that do not recognize an IBB domain, such as transportin³⁷ and exportin-tRNA^{38,39}. The distance between the acidic loop and HR1 is such that Ran could feasibly contact both simultaneously. Ran has a C-terminal acidic tail containing a conserved DEDDDL motif, which, in the structure of Ran-GDP, is located close to a basic patch on the protein surface⁴⁰. This acidic tail is sequestered by the Ran-binding domain, RanBD1, in the structure of RanBD1 bound to a Ran-GTP-analogue complex⁴¹. Furthermore, the N terminus of RanBD1 also contains a DDDDD motif which, in the complex, is located close to the basic patch of Ran. Thus, the acidic RanBD1 N terminus appears to take the place of the acidic Ran C terminus upon complex formation. The conserved acidic loop of importin- β may interact with Ran in a similar way. Because Trp 342 in this loop is involved in the recognition of the N-terminal IBB moiety, it is possible that engagement of the loop by Ran-GTP would promote dissociation of the IBB domain.

Importins α and β compared

The structure of importin- β in complex with the IBB domain is very similar to that of importin- α bound to an NLS peptide¹⁸. A good alignment (r.m.s.d._{95Ca} = 1.47 Å) can be made between HEAT repeats 9–12 and armadillo repeats 2–5, regions that interact with the N-terminal moiety of the IBB domain and with the NLS peptide, respectively (Fig. 5). The alignment becomes poorer when extended to neighbouring repeats, because although both proteins form a right-handed superhelix, importin- β is more tightly wound (the pitch and mean diameter are both ~ 50 Å for importin- β , compared with 100 Å and 30 Å for importin- α). The alignment places the IBB N-terminal moiety and the NLS peptide in similar locations, but with opposite chain directions. The nature of the binding interaction is similar in both cases: conserved acidic residues of importin- α or - β interact with basic residues of the IBB or NLS substrate, whose side chains are positioned by tryptophan or other hydrophobic residues. In the alignment of Fig. 5, two such tryptophans occur at the same position for each importin: Trp 430 in HEAT repeat 10 matches α Trp 195 in arm-3, and Trp 472 in HR11 matches α Trp 237 in arm-4. All four tryptophan residues interact with lysines of the substrates, and the superposition places the two lysines of the IBB domain very close to the two of the NLS (corresponding N γ atoms are either 2.8 Å or 1.7 Å apart). The similar backbone fold and spatial arrangement of interacting elements strongly support the hypothesis that importins- α and - β evolved from a common ancestor³⁴, and more generally that all HEAT- and armadillo-repeat-containing proteins belong to the same protein superfamily.

The functional similarity of the two importins is underscored by the fact that certain proteins are imported into the nucleus bound directly to importin- β without first binding to importin- α . One such protein is HIV-1 Rev. Rev interacts with importin- β through an arginine-rich region which is also involved in binding RNA³². The N-terminal 58 residues of Rev containing this region can be reasonably aligned with the sequence of the IBB domain (not shown), suggesting that the N terminus of Rev may form an IBB-like structure. In support of this hypothesis, the nuclear magnetic resonance structure of a Rev peptide bound to an RNA Rev-response element shows that residues 34–50, which align with residues in the C-terminal IBB helix, do indeed form a helix⁴². Whether Rev mimics an IBB domain to penetrate the nucleus and whether other proteins employ a similar strategy are questions for future study. □

Methods

Protein expression, purification and crystallization. *Escherichia coli* strain M15 cells expressing human importin- β from a pQE-30 expression vector

(QIAGEN) were grown in LB medium until attaining an optical density of 0.5 at 600 nm, induced with 0.5 mM isopropylthiogalactoside and grown for a further 5 h at 30 °C before harvesting. For selenomethionine incorporation, cells were grown in minimal medium M9 and supplemented with selenomethionine and other essential amino acids just before induction. We purified importin- β using an importin- α affinity column essentially as described²⁰. The IBB domain comprising residues α 11– α 54 of importin- α was produced by solid-phase synthesis and purified by reversed-phase high performance liquid chromatography. Importin- β at a concentration of 50 mg ml⁻¹ was incubated in a 1:3 molar ratio with the IBB domain at 20 °C, and the complex was crystallized by either hanging drop vapour diffusion (crystal form I) or the batch method (form II). Crystal form I grew from 18–22% PEG8000, 50 mM potassium acetate pH 5.8, and 10 mM β -mercaptoethanol. Crystals of form II grew from 16–20% PEG8000, 50 mM potassium phosphate pH 3.9, 10 mM β -mercaptoethanol and 5 mM sodium iodide, and were allowed to become partially dehydrated for 1–7 days in a hanging drop well which contained no reservoir solution. Crystals were harvested in their mother liquor and flash-frozen in liquid nitrogen using 25% PEG400 as the cryoprotectant.

Data collection and structure determination. Native and MAD diffraction data for crystal form I were collected on a MarResearch image plate and a Mar CCD detector, respectively (Table 1). Data were reduced using the HKL package⁴³ and further processed with programs from the CCP4 suite⁴⁴. Twenty-two selenium atoms (out of a possible 24) were successfully located using the program SOLVE⁴⁵, and the initial electron-density map showed a distinct solvent boundary and regions of helical density. A map calculated at 5 Å revealed the overall protein fold and permitted a HEAT-repeat model fashioned from that of the protein phosphatase 2A PR65/A subunit to be crudely positioned as a guide for tracing. Thirty-eight polyaniline helices were placed into the 3.3 Å experimental density using program O⁴⁶ and subjected to rigid body refinement using program CNS⁴⁷. Recombination of MAD phases with those calculated from the model yielded a map allowing much of the sequence to be traced. The chain trace of importin- β was greatly expedited by the selenium peaks in an anomalous difference Fourier. Tracing of the IBB domain was facilitated through use of a mercurated IBB domain, in which α Ala 12 was replaced by a cysteine during the peptide synthesis step and subsequently derivatized with p-chloromercuribenzoate. A difference Fourier using data from a crystal grown with the mercurated IBB domain revealed an 8 σ Hg peak next to residue α 12. Refinement was carried out in CNS by a simulated annealing, torsion-angle refinement protocol which minimized the maximum-likelihood 'MLHL' target function with the MAD phases serving as additional restraints. Once R_{free} had dropped to 27.9%, peaks above 3 σ in the $F_o - F_c$ difference Fourier and within hydrogen-bonding distance to protein atoms were modelled as water molecules, leading to a further drop in R_{free} . The final model has R_{cryst} = 22.1% and R_{free} = 27.3%, and consists of IBB residues α 11– α 54, all 876 importin- β residues, and 44 water molecules. The model has good geometry (see Table 1), with 90.8% residues in the most favoured regions of the Ramachandran plot, and only residue α Lys 18 in a disallowed region (see text).

We collected native diffraction data for crystal form II on a Mar CCD detector. Weak diffraction could be observed at up to 2.05 Å resolution, but the ratio $\langle I/\sigma I \rangle$ drops below 2.0 in the 2.15–2.20 Å shell. The structure was solved by molecular replacement with program AMORE⁴⁴ using the refined model from crystal form I as the search model. The correct solution corresponded to the top peak in both the rotation and the translation function. After rigid-body refinement of individual helices, a 2 $F_o - F_c$ map revealed residues 1–610 to be in good density apart from a few loops. In contrast, substantial rebuilding was required for residues 611–876 and in the IBB domain. The final refined model has good geometry (Table 1), with no residues in disallowed regions of the Ramachandran plot.

Received 13 April; accepted 27 April 1999.

- Görlich, D. & Mattaj, J. W. Nucleocytoplasmic transport. *Science* **271**, 1513–1518 (1996).
- Mattaj, J. W. & Englmeier, L. Nucleocytoplasmic transport: the soluble phase. *Annu. Rev. Biochem.* **67**, 265–306 (1998).
- Weis, K. Importins and exportins: how to get in and out of the nucleus. *Trends Biochem. Sci.* **23**, 185–189 (1998).
- Doye, V. & Hurt, E. From nucleoporins to nuclear pore complexes. *Curr. Opin. Cell Biol.* **9**, 401–411 (1997).

5. Ohno, M., Fornerod, M. & Mattaj, I. W. Nucleocytoplasmic transport: the last 200 nanometers. *Cell* **92**, 327–336 (1998).
6. Dingwall, C. & Laskey, R. A. Nuclear targeting sequences: a consensus? *Trends Biochem. Sci.* **16**, 178–181 (1991).
7. Kalderon, D., Roberts, B. L., Richardson, W. D. & Smith, A. E. A short amino acid sequence able to specify nuclear location. *Cell* **39**, 499–509 (1984).
8. Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. Two interdependent basic domains in nucleoplasmin targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615–623 (1991).
9. Görlich, D., Prehn, S., Laskey, R. A. & Hartmann, E. Isolation of a protein that is essential for the first step of nuclear import. *Cell* **79**, 767–778 (1994).
10. Weis, K., Mattaj, I. W. & Lamond, A. I. Identification of hSRP1 α as a functional receptor for nuclear localization sequences. *Science* **268**, 1049–1053 (1995).
11. Moroianu, J., Blobel, G. & Radu, A. Previously identified protein of uncertain function is karyopherin α and together with karyopherin β docks import substrate at nuclear pore complex. *Proc. Natl Acad. Sci. USA* **92**, 2008–2011 (1995).
12. Görlich, D. *et al.* Two different subunits of importin cooperate to recognize nuclear localization signals and bind them to the nuclear envelope. *Curr. Biol.* **5**, 383–392 (1995).
13. Chi, N. C., Adam, E. J. H. & Adam, S. A. Sequence and characterization of cytoplasmic nuclear import factor p97. *J. Cell Biol.* **130**, 265–274 (1995).
14. Imamoto, N. *et al.* The nuclear pore-targeting complex binds to nuclear pores after association with a karyophile. *FEBS Lett.* **368**, 415–419 (1995).
15. Radu, A., Blobel, G. & Moore, M. S. Identification of a protein complex that is required for nuclear import and mediates docking of import substrates to distinct nucleoporins. *Proc. Natl Acad. Sci. USA* **92**, 1769–1773 (1995).
16. Izaurralde, E., Kutay, U., von Kobbe, C., Mattaj, I. W. & Görlich, D. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *EMBO J.* **16**, 6535–6547 (1997).
17. Bischoff, F. R. & Görlich, D. RanBP1 is crucial for the release of RanGTP from importin β -related nuclear transport factors. *FEBS Lett.* **419**, 249–254 (1997).
18. Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α . *Cell* **94**, 193–204 (1998).
19. Görlich, D., Henklein, P., Laskey, R. A. & Hartmann, E. A 41 amino acid motif in importin- α confers binding to importin- β and hence transit into the nucleus. *EMBO J.* **15**, 1810–1817 (1996).
20. Weis, K., Ryder, U. & Lamond, A. I. The conserved amino-terminal domain of hSRP1 α is essential for nuclear import. *EMBO J.* **15**, 1818–1825 (1996).
21. Kobe, B. Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin α . *Nature Struct. Biol.* **6**, 388–397 (1999).
22. Kutay, U., Izaurralde, E., Bischoff, F. R., Mattaj, I. W. & Görlich, D. Dominant-negative mutants of importin- β block multiple pathways of import and export through the nuclear pore complex. *EMBO J.* **16**, 1153–1163 (1997).
23. Chi, N. C. & Adam, S. A. Functional domains in nuclear import factor p97 for binding the nuclear localization sequence receptor and the nuclear pore. *Mol. Biol. Cell* **8**, 945–956 (1997).
24. Chi, N. C., Adam, E. J. H. & Adam, S. A. Different binding domains for Ran-GTP and Ran-GDP/RanBP1 on nuclear import factor p97. *J. Biol. Chem.* **272**, 6818–6822 (1997).
25. Görlich, D. *et al.* A novel class of RanGTP binding proteins. *J. Cell Biol.* **138**, 65–80 (1997).
26. Andrade, M. A. & Bork, P. HEAT repeats in the Huntington's disease protein. *Nature Genet.* **11**, 115–116 (1995).
27. Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A. & Barford, D. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs. *Cell* **96**, 99–110 (1999).
28. Palacios, I., Hetzer, M., Adam, S. A. & Mattaj, I. W. Nuclear import of U snRNPs requires importin β . *EMBO J.* **16**, 6783–6792 (1997).
29. Huber, J. *et al.* Snurportin1, an m⁷G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.* **17**, 4114–4126 (1998).
30. Jäkel, S. & Görlich, D. Importin β , transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J.* **17**, 4491–4502 (1998).
31. Moore, J. D., Yang, J. Y., Truant, R. & Kornbluth, S. Nuclear import of Cdk/cyclin complexes: Identification of distinct mechanisms for import of Cdk2/cyclin E and Cdc2/cyclin B1. *J. Cell Biol.* **144**, 213–224 (1999).
32. Truant, R. & Cullen, B. R. The arginine-rich domains present in human immunodeficiency virus type 1 Tat and Rev function as direct importin β -dependent nuclear localization signals. *Mol. Cell Biol.* **19**, 1210–1217 (1999).
33. Palmeri, D. & Malim, M. H. Importin β can mediate the nuclear import of an arginine-rich nuclear localization signal in the absence of importin α . *Mol. Cell Biol.* **19**, 1218–1225 (1999).
34. Malik, H. S., Eickbush, T. H. & Goldfarb, D. S. Evolutionary specialization of the nuclear targeting apparatus. *Proc. Natl Acad. Sci. USA* **97**, 13738–13742 (1997).
35. Esnouf, R. M. *et al.* Continuous and discontinuous changes in the unit cell of HIV-1 reverse transcriptase crystals on dehydration. *Acta Crystallogr. D* **54**, 938–953 (1998).
36. Hayward, S. & Berendsen, H. J. C. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase and T4 lysozyme. *Proteins* **30**, 144–154 (1996).
37. Pollard, V. W. *et al.* A novel receptor-mediated nuclear protein import pathway. *Cell* **86**, 985–994 (1996).
38. Kutay, U. *et al.* Identification of a tRNA specific nuclear export receptor. *Mol. Cell* **1**, 359–369 (1998).
39. Arts, G.-J., Fornerod, M. & Mattaj, I. W. Identification of a nuclear export receptor for tRNA. *Curr. Biol.* **8**, 305–314 (1998).
40. Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. & Wittinghofer, A. Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. *Nature* **374**, 378–381 (1995).
41. Vetter, I. R., Nowak, C., Nishimoto, T., Köhlmann, J. & Wittinghofer, A. Structure of a Ran-binding domain complexes with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39–46 (1999).
42. Battiste, J. L. *et al.* α -Helix-RNA major groove recognition in an HIV-1 Rev peptide-RRE RNA complex. *Science* **273**, 1547–1551 (1996).
43. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
44. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
45. Terwilliger, T. C., Kim, S.-H. & Eisenberg, D. Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1–5 (1987).
46. Jones, T. A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
47. Brünger, A. T. *et al.* Crystallography and NMR system: A new software for macromolecular structure determination. *Acta Crystallogr. C* **54**, 905–921 (1998).
48. Kraulis, P. E. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
49. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insight from the interfacial and thermodynamic properties of hydrocarbon. *Proteins* **11**, 281–296 (1991).
50. Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).

Acknowledgements. We thank M. Moulin for excellent technical assistance; R. W. Frank at the Zentrum für Molekulare Biologie, Universität Heidelberg for peptide synthesis; members of the EMBL/ESRF Joint Structural Biology Group, in particular G. Leonard and A. Thompson for access and support at beamline BM14, W. Burmeister at beamline ID14-3 and J. Lescar and B. Rasmussen at beamline ID02; and D. Barford for providing us with the coordinates of the PR65/A subunit of PP2A before release. We acknowledge our use of the HKL package as part of a collaboration with Z. Otwinowski and W. Minor, supported by the NIH. C.P. was supported by an EMBO fellowship and by a Marie Curie (TMR) fellowship.

Correspondence and requests for materials should be addressed to C.W.M. (e-mail: mueller@embl-grenoble.fr). Atomic coordinates of crystal forms I and II have been deposited with the PDB under access codes 1QGG and 1QGR, respectively.

Structure of the nuclear transport complex karyopherin- β 2-Ran-GppNHp

Yuh Min Chook & Günter Blobel

Laboratory of Cell Biology, Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10021, USA

Transport factors in the karyopherin- β (also called importin- β) family mediate the movement of macromolecules in nuclear-cytoplasmic transport pathways. Karyopherin- β 2 (transportin) binds a cognate import substrate and targets it to the nuclear pore complex. In the nucleus, Ran-GTP binds karyopherin- β 2 and dissociates the substrate. Here we present the 3.0 Å structure of the karyopherin- β 2-Ran-GppNHp complex where GppNHp is a non-hydrolysable GTP analogue. Karyopherin- β 2 contains eighteen HEAT repeats arranged into two continuous orthogonal arches. Ran is clamped in the amino-terminal arch and substrate-binding activity is mapped to the carboxy-terminal arch. A large loop in HEAT repeat 7 spans both arches. Interactions of the loop with Ran and the C-terminal arch implicate it in GTPase-mediated dissociation of the import-substrate. Ran-GppNHp in the complex shows extensive structural rearrangement, compared to Ran-GDP, in regions contacting karyopherin- β 2. This provides a structural basis for the specificity of the karyopherin- β family for the GTP-bound state of Ran, as well as a rationale for interactions of the karyopherin-Ran complex with the regulatory proteins ranGAP, ranGEF and ranBP1.

Bidirectional nuclear-cytoplasmic transport of macromolecules through nuclear pore complexes (NPC) is mediated by transport factors of the karyopherin- β (Kap- β) family. Kap- β s target transport substrates to the NPC by interacting with both specific NPC proteins (nucleoporins) and a cognate nuclear-localization sequence (NLS) or nuclear-export sequence (NES) on import or export substrates, respectively¹. Subsequent translocation through the NPC occurs in the presence of the GTPase Ran and several Ran-binding proteins such as ranBP1 (ref. 1) and p10 (refs 2, 3). Various substrates are transported into the nucleus by several distinct pathways, all mediated by different members of the Kap- β family. The first (or 'classical') transport pathway involves import of proteins with a 'classical' NLS, characterized by one or two stretches of basic residues, by the Kap- α /Kap- β 1 heterodimer. Kap- β 1 binds the adaptor protein Kap- α , which in turn binds the NLS⁴. Other transport pathways have since emerged, all involving transport factors with limited sequence identity to Kap- β 1. These include import pathways for messenger RNA-binding proteins using Kap- β 2, Kap104 and Kap111/Mtr10 (ref. 1); ribosomal subunits using Kap- β 3, Kap121 and Kap123 (ref. 1); RNA chaperone Lhp1p using Kap108/Sxm11 (ref. 1); and transcription factors TFIIS and Pho4 using Kap119/Nmd5 (ref. 5) and Kap121 (ref. 6), respectively. Export pathways include the export of proteins bearing classical leucine-rich NES by Crm1/Kap124 (ref. 1); Kap- α export by Cas/Kap109 (ref. 1); transfer RNA export by hLos1/exportin-t (ref. 1); and Pho4 export by Msn5 (ref. 7). All of these Kap- β s bind substrates directly through specific NLS/NES.

In all identified transport systems, Kap- β s bind Ran-GTP. In

import systems, the formation of complexes between GTPase and Kap- β s is involved in substrate dissociation in the nucleus^{5,8–10}. In contrast, interactions between karyopherin, Ran-GTP and substrates appear to be cooperative in export systems¹. The nucleotide state of Ran is regulated by ranGAP and ranGEF, the former catalysing nucleotide hydrolysis to form Ran-GDP and the latter catalysing nucleotide exchange to regenerate Ran-GTP¹¹. Current models postulate that nuclear localization of ranGEF and the resulting nuclear Ran-GTP are responsible for the final step of nuclear import, the release of substrate from Kap- β s^{8,10,11}. The complexing of Kap- β s and Ran-GTP may also constitute the first step of export: either recycling of Kap- β s to the cytoplasm following completion of import¹⁰ or as a ternary complex with an export substrate¹. Despite the identification of components of the different transport pathways, the molecular mechanism of nuclear transport remains unclear. Key mechanistic issues include GTPase regulation of Kap- β -substrate complexes: how does Ran-GTP dissociate Kap- β -import substrate complexes, and how does Ran-GTP binding to certain Kap- β s promote binding of export substrates? The issue of substrate recognition, such as identification of non-classical NLS/NES and elements in Kap- β s important for NLS/NES recognition, also remains to be explored. Finally, little is known about the mechanism of movement of transport complexes through the NPC following initial docking.

To gain insight into the molecular details of nuclear transport, we have focused on a prototypic import pathway, that of Kap- β 2, which mediates nuclear import of the pre-mRNA-binding proteins A1 (refs 12–14) and F (ref. 9). Kap- β 2 (or transportin) recognizes

Table 1 Crystallographic analysis

Data collection statistics	Wavelength (Å)	dmin (Å)	Completeness (%)	Observed reflections	Unique reflections	R sym*
Crystal						
Semet Kap- β 2-Ran-GppNHp	0.9800	3.0	99.4	331,713	54,577	0.095
	0.9798	3.0	98.9	379,166	54,341	0.107
	0.9500	3.4	99.4	227,870	37,528	0.105
Refinement statistics						
R-factor† (%)	26.7			R.m.s.d. from ideal bond lengths	0.009 Å	
R _{free} † (%)	31.5			R.m.s.d. from ideal bond angles	1.46°	
Number of atoms	8,571					

* Rsym = $\sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)$; $I_i(h)$ is the i th measurement of reflection h and $\langle I(h) \rangle$ is the weighted mean of all measurements of h .

† R-factor = $\sum_h ||F_{obs}(h)| - |F_{calc}(h)|| / \sum_h |F_{obs}(h)|$; R_{free} is calculated with 5% of the data.

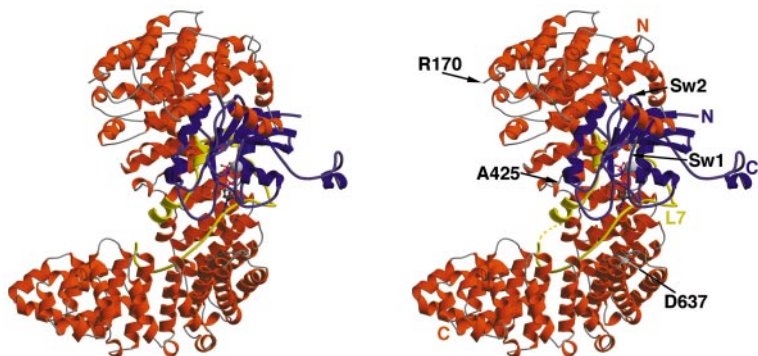


Figure 1 Stereoview of a ribbon drawing of Kap-β2 bound to Ran-GppNHp. Kap-β2 is orange and residues 311–373 (L7) are highlighted in yellow to emphasize its deviation from the HEAT repeat structure and its displacement from the main

body of the protein. Ran is blue; the nucleotide analogue GppNHp is depicted with ball-and-stick representation and Mg²⁺ is a white sphere. Ribbon diagrams were generated using MOLSCRIPT⁴⁸ and RASTER3D⁴⁹.

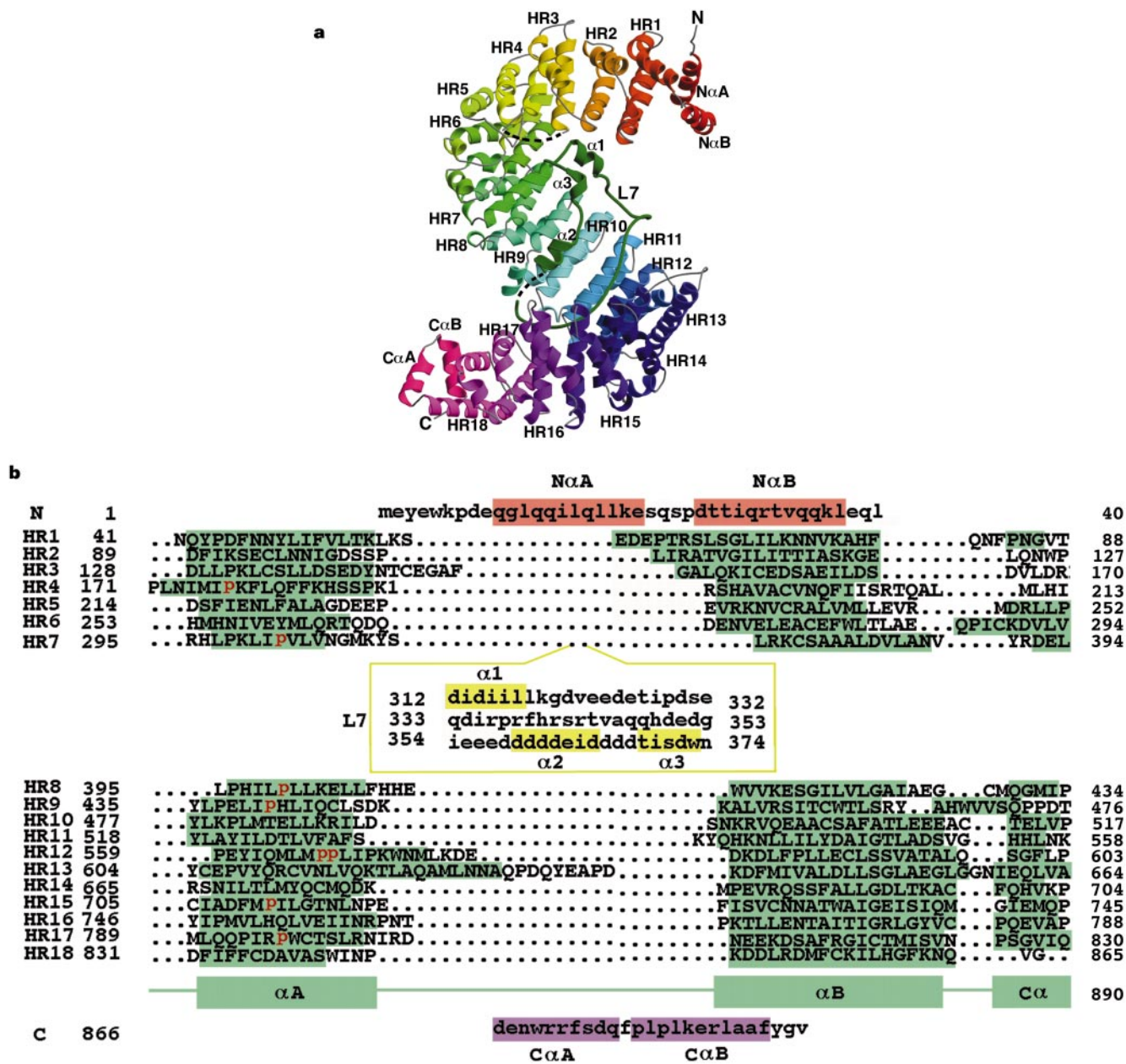


Figure 2 Kap-β2 contains eighteen HEAT repeats. **a**, Ribbon diagram of Kap-β2 with each HEAT repeat in a different colour and labelled HR1–HR18. The two helices (NαA, NαB) N-terminal to HR1 are red and the helices that cap the C terminus (CαA, CαB) are pink. Loop L7, which connects helices in HR7 with short α-helical regions α1, α2 and α3, is dark green. **b**, Sequence of Kap-β2 and

the corresponding secondary structure. Residues in helical conformation are shaded. HEAT repeats HR1–HR18 are grouped such that residues at equivalent positions in helices αA, αB and αC are aligned. The resulting HEAT repeats alignment is strictly structural. Proline residues in the middle of helices αA are red.

the 39-residue Gly/Asn-rich NLS in A1 termed M9 (refs 12, 13). Kap- β 2 binds Ran-GTP and Kap- β 2-A1 is dissociated by the GTPase^{9,10}. We present here the 3.0 Å crystal structure of the complex of human Kap- β 2-Ran-GppNHp (a non-hydrolysable GTP analogue), which has a relative molecular mass of 125,000 ($M_r = 125K$). Kap- β 2 has 18 tandem HEAT repeats (named after the proteins huntingtin, elongation factor 3, 'A' subunit of protein phosphatase 2A, and TOR1 lipid kinase, where the motif was first identified)¹⁵ arranged into a right-handed spiral of two orthogonal arches. Ran interacts extensively with Kap- β 2 in the N-terminal arch, and deletion mutants^{13,14} map A1 binding to the other arch. The structure indicates that Ran-mediated structural changes in a long loop in HEAT repeat 7, which extends into the substrate-binding arch, may account for substrate dissociation by the GTPase. Ran-GppNHp in the Kap- β 2 complex shows extensive structural rearrangement when compared to Ran-GDP^{16,17}. The regions of Ran affected by the nucleotide switch are also part of the Kap- β 2-Ran interface, thus providing a rationale for the specificity of Kap- β s for the GTP state of Ran. The structure of the complex also provides insight into higher-order interactions with regulators and effectors of Ran such as ranGAP, ranGEF, ranBP1 and ranBP1-homology (RBH) domains of the nucleoporin Nup358.

Overall structure

Crystals of the human Kap- β 2-Ran-GppNHp complex diffract to 3.0 Å, and are of the space group $P3_121$ ($a = b = 133.11$ Å, $c = 138.41$ Å). They consist of one molecule of selenomethionine-labelled Kap- β 2 complexed to selenomethionine-labelled Ran-GppNHp in each asymmetric unit. The structure was solved using multiwavelength anomalous dispersion (MAD; Table 1). We determined positions of 24 well-ordered selenium atoms (of 30 possible selenomethionines) using SHELX-97 Patterson interpretation¹⁸ and difference Fourier methods. The final model has an R -value of 26.7% (free R -value 31.6%; Table 1) and includes 880 of the 890 residues of Kap- β 2 (3–166, 170–352 and 358–890), 190 of the 216 residues of Ran (8–197), GppNHp and a magnesium ion.

Kap- β 2 is exclusively helical and contains eighteen HEAT repeats (Figs 1, 2a). Despite the large size of Kap- β 2 ($M_r = 101.3K$), the helical repeats produce a single contiguous hydrophobic core, explaining the generally poor biochemical behaviour of truncated fragments of the protein and our unsuccessful attempts to obtain crystals thereof. This property of the Kap- β fold indicates that

studies employing truncated mutants of proteins in the family should be interpreted with caution. The HEAT repeats of Kap- β 2 are organized in a right-handed superhelix to produce a highly twisted structure of two continuous orthogonal arches (Fig. 1). The orthogonal arrangement of the arches results in a compact structure, with a distance of 120 Å between the N and C termini. The N-terminal arch is approximately 90 Å wide and 70 Å high (outer dimensions) and spans residues 1–638, whereas the C-terminal arch is more compact (65 Å wide and 55 Å high), and spans residues 411–890 (Fig. 1). There is a long loop connecting the helices of HEAT repeat 7 (L7, residues 311–373; Fig. 2a, b). The proximal portion of L7 lines the last third of the N-terminal arch and the distal portion is in the C-terminal arch (Figs 1, 2a). Ran-GppNHp is clamped into the N-terminal arch of Kap- β 2 through two distinct contact sites. The N-terminal helix bundle ($N\alpha A$, $N\alpha B$) and HEAT repeats 1 and 2 of Kap- β 2 pack against switch 2 and part of switch 1 of Ran (Fig. 1). Also, the proximal and central portions of the L7 loop pack against the $\alpha 4$ and $\alpha 5$ helices and the $\beta 6$ strand of Ran (residues 124–173; Fig. 1).

Organization of Kap- β 2 HEAT repeats

The HEAT repeats of Kap- β 2 resemble those in the A subunit of protein phosphatase 2A (ref. 19) and the armadillo repeats in Kap- α (ref. 4) and β -catenin²⁰. These repeat-containing proteins have similar overall protein architecture, with units of two antiparallel helices stacked consecutively to form two layers of parallel helices. The Kap- β 2 N- and C-terminal three helices form triangular helical arrays to cap both termini (Fig. 2a). The first and last two helices ($N\alpha A$, $N\alpha B$ and $C\alpha A$, $C\alpha B$) are not parallel but are almost perpendicular to adjacent helices and thus not defined as HEAT repeats. This definition is also consistent with a previous prediction of HEAT-repeat distribution in Kap- β 2 (ref. 21). The first HEAT repeat in our structure begins at the third helix (HR1 αA), as this is the first helix that stacks parallel to an adjacent helix (HR2 αA).

Most of the Kap- β 2 HEAT repeats (HR1, HR5–HR17) are composed of three α -helices (αA , αB and αC) reminiscent of 3-helix armadillo repeats^{4,20} (Fig. 2a, b). Within each HEAT repeat, helices αA and αB are antiparallel, contain 12–24 residues each, and are followed by helix αC (3–9 residues) or by a loop. Equivalent helices in the HEAT repeats are parallel, with αBs lining the concave surface of the protein and αAs forming the spine or convex side of the arches. In HR2, HR3 and HR18, αC is replaced by a loop; in HR4, residues

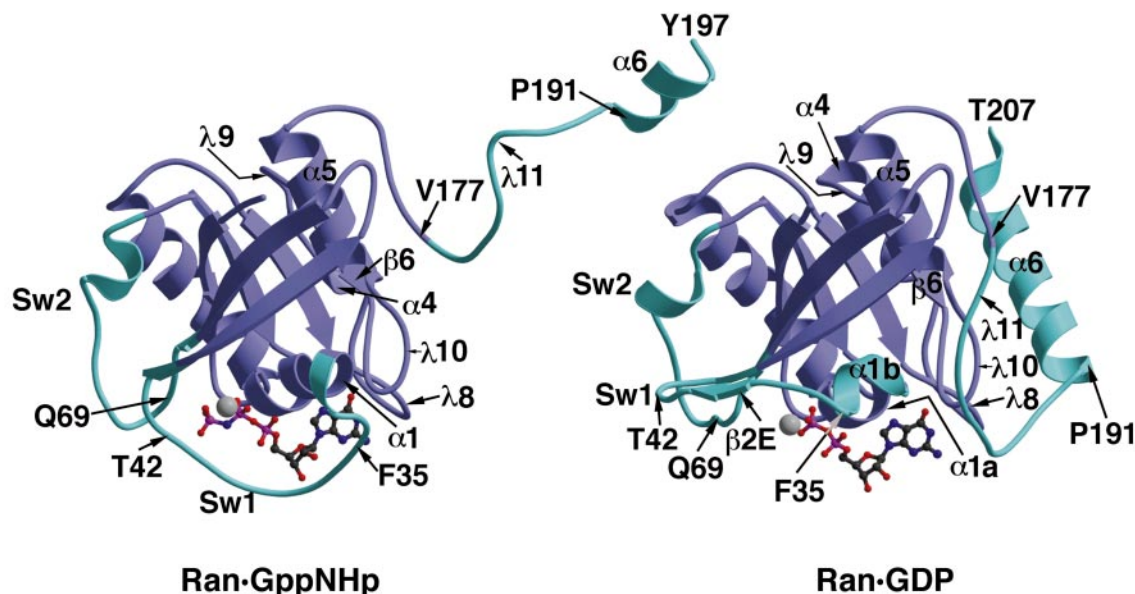


Figure 3 Structural differences between Ran-GppNHp and Ran-GDP. Regions of Ran that are different in the GppNHp versus GDP states are in cyan. These regions include residues spanning helix $\alpha 1$ to $\beta 2$, also known as switch 1 (Sw1),

switch 2 (Sw2) and the C-terminal extension (residues 176–207). Coordinates of Ran-GDP were obtained from the Protein Data Bank, accession code 1byu¹⁶.

connecting HR4 α B and HR5 α A are disordered. These 2-helix repeats resemble the 2-helix HEAT repeats in protein phosphatase 2A (ref. 19). Another common feature in the Kap- β 2 and phosphatase 2A HEAT repeats is the presence of prolines in the middle of many outer helices (α A), which results in a bend in these elements. HR4, HR7–HR9, HR15 and HR17 each has a proline in the second turn of α A, and HR12 has two consecutive prolines in the third turn of its α A (Fig. 2b).

The relative orientation of neighbouring HEAT repeats determines the overall conformation of Kap- β 2. Rotations of about 15° between adjacent repeat units alternate with four sets of 30°–60° rotations, bending the HEAT repeat progression four times to form the two arches. Large rotations relate HR2–HR3 (40°), HR3–HR4 (35°) and HR5–HR6 (40°) at the first bend and the HR8–HR9 30° rotation results in the second bend, thus forming the N-terminal arch. The next two bends are clustered at HR12–HR13 (40°) and HR13–HR14 (40°), then at HR15–HR16 (43°), HR16–HR17 (59°) and HR17–HR18 (50°) to form the more compact C-terminal arch.

Ran nucleotide switch

The affinity of Kap- β s for Ran-GTP is about 10⁴-fold higher than for Ran-GDP²². Therefore, understanding the structural basis of nucleotide switching will help us to understand the specificity of Kap- β s for the GTP state of Ran. The nucleotide-switch regions (switch 1: residues 30–47; switch 2: residues 65–80) of Ran-GDP differ from those of most members of the Ras superfamily¹⁷, with a different chain direction in helix α 1B and an additional β -strand (β 2E; Fig. 3). Furthermore, the Ran sequence contains a 40-residue C-terminal extension (residues 177–216) not present in other Ras superfamily members. In Ran-GDP, this sequence forms extended chain λ 11 and helix α 6, which pack against the G-domain of the protein (Fig. 3)¹⁷. Ran-GppNHp in the Kap- β 2 complex shows large structural rearrangements when compared to Ran-GDP. These changes are localized to the switch regions and the C-terminal extension, whereas the G-domains are similar (root-mean-square deviation (r.m.s.d.) 0.5 Å, superposition of the central β -sheets;

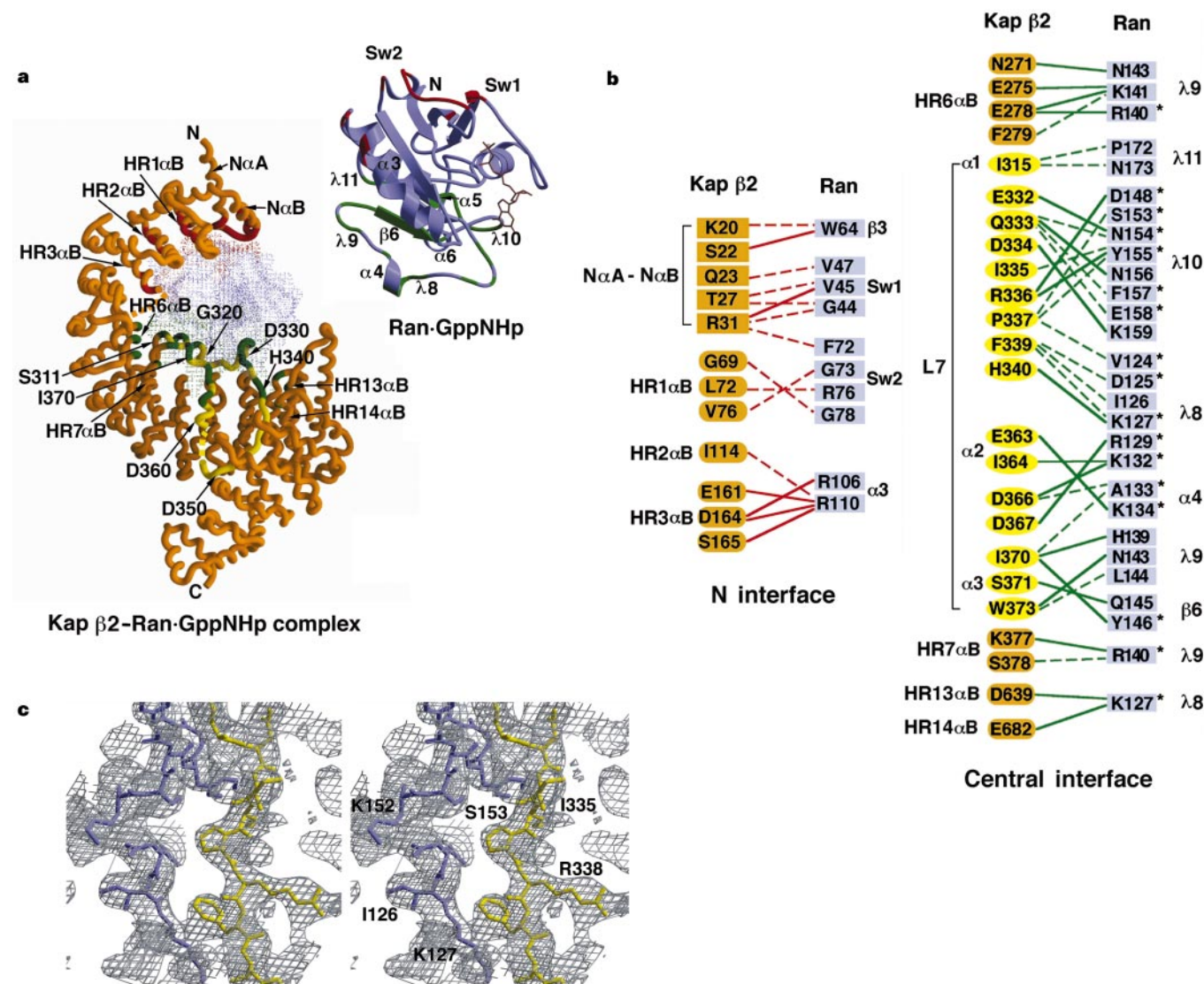


Figure 4 Kap- β 2-Ran-GppNHp interface. **a**, The Kap- β 2-Ran-GppNHp complex showing two distinct interfaces (red and green, respectively) in both proteins. Kap- β 2 is shown as an orange backbone tube with yellow L7 and Ran is depicted as a blue dot surface. In both proteins, residues at the N interface and the central interface are red and green, respectively. Every tenth residue in L7 is labelled. Ribbon diagram of Ran-GppNHp is oriented and coloured as in the Ran dot surface representation in the complex. **b**, Schematic representation of interactions at the N interface and the central interface. Ran-GppNHp residues are blue and Kap- β 2 residues are orange, except for L7 residues 311–373 (yellow). At the N

interface, van der Waals contacts (<4.0 Å) are represented by red dashed lines and polar contacts by solid red lines. At the central interface, van der Waals contact (<4.0 Å) are represented by green dashed lines and polar contacts are represented by solid green lines. All polar interactions shown are <3.5 Å except for charged contacts D164–R106 and D164–R110. Residues in Ran-GDP that are buried by the C-terminal extension¹⁶ are asterisked. **c**, MAD-phased, density-modified electron density coloured at 1.0 σ at the central-interface of the Kap- β 2-Ran-GppNHp complex. Ran residues are blue and Kap- β 2 residues yellow.

Fig. 3). The structure of Ran-GppNHp bound to a Ran binding domain (RBD1) appears to be similar to the GTPase in our Kap- β 2 complex in the switch regions, but conformations of the C-terminal extensions vary significantly in the two complexes²³.

Although Ran-GDP deviates from the stereotypical Ras fold²⁴, the triphosphate state in our complex is highly similar in overall structure and details of Mg²⁺ and nucleotide coordination (Fig. 3). In our structure, the nucleotide γ -phosphate contacts main-chain amides of T42 and G68 (Ras T35, G60), and F35 (F28 in Ras) forms hydrophobic interactions with the nucleotide base, as in Ras²⁴. Furthermore, although Mg²⁺ is coordinated in the GDP state by switch 2 residues T66 and E70 (ref. 16), in Ran-GppNHp, as in Ras, Mg²⁺ is coordinated by the side chains of T42 and D65 (Ras D57). These changes require a marked conformational rearrangement of the Ran-GDP structure. Upon binding the GTP analogue, helix α 1a extends from D27 to L31, helix α 1b dissolves into an extended chain and a peptide bond flip at L31 redirects the chain so that the overall structure of switch 1 is similar to that in Ras (Fig. 3). The similarity to Ras extends into switch 2 or Ran-GppNHp, where the Q69 side chain (Q61 in Ras), crucial in nucleotide hydrolysis, has moved towards the γ -phosphate. Similarity of the Ran switch regions to Ras²⁴, as well as to Ran in the RBD1 complex²³, indicate that the structural differences seen in these regions when compared to Ran-GDP are due to nucleotide switching rather than Kap- β 2 complexation.

The structural rearrangements in switch 1 are coupled to those in the C-terminal extension of Ran. In Ran-GDP¹⁶, helices α 1a,b are held close to residues 180–186 in the C-terminal extension by interactions that include main-chain hydrogen bonds between R29 and A183 and between L31 and A181 (Fig. 3). In the GTP state, these hydrogen bonds are disrupted and replaced with intra-helix hydrogen bonds as helix α 1 extends to L31. Residues G33 and E34 move into the space previously occupied by A183 in the GDP state, displacing the extended chain λ 11 of the C-terminal extension. A similar displacement of λ 11 occurs in the Ran-GppNHp–RBD1 structure and the conformation of the C-terminal extension is altered as it wraps around RBD1 (ref. 23). In the Kap- β 2–Ran-GppNHp complex, the C-terminal extension of Ran (residues 177–197) is extruded from the body of the GTPase and residues 182–193 are involved in crystal contacts with Kap- β 2 in another asymmetric unit (Fig. 3). Based on the structure alone, it is not certain whether extrusion is a result of nucleotide switch, Kap- β 2 complexation, or both. Monoclonal antibodies recognizing residues 210–216 of Ran bind Kap- β –Ran-GTP/Ran-GDP complexes but not the free GTPases, indicating that the C terminus of Ran is exposed only in complex with Kap- β s²⁵. However, antibodies against adjacent residues 196–207 preferentially bind the triphosphate state of the free GTPase, indicating that the C-terminal extension of Ran-GTP is more weakly associated with the G-domain and is therefore more accessible than in Ran-GDP²². These studies and our structural observations indicate that nucleotide-switch-induced structural rearrangement of switch 1 and displacement of λ 11 may facilitate the detachment of a portion of the C-terminal extension, but its complete extrusion is probably due to complexing with Kap- β 2.

Kap- β 2–Ran interface

The formation of the Kap- β 2–Ran-GppNHp complex buries 3,900 Å² of accessible surface area in the two proteins. Two distinct interfaces are found in the complex (Fig. 4). The first interface (N interface) involves interactions between helix N α B, helices α B of HR1–HR3 in the N-terminal region of Kap- β 2 and the switch regions of Ran. The second (central interface) involves loop L7 and, to a lesser degree, the α B helices of HR6, HR7, HR13 and HR14 in the central region of Kap- β 2 interacting with the face of Ran, composed of helices α 4 and α 5, strand β 6 and intervening loops. The two interfaces are spatially distinct in Kap- β 2, as HR4 and HR5 do not contact Ran.

Residues buried at the N interface are highlighted in Fig. 4a and

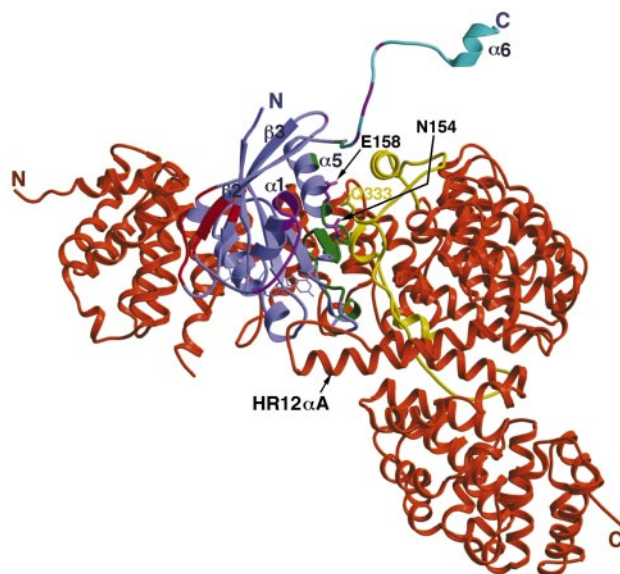


Figure 5 Kap- β 2–Ran and RBD1–Ran interfaces are mostly non-overlapping but adjacent in the Ran structure. RBD1 may be able to access Ran through the opening of the N-terminal arch. Kap- β 2–Ran-GppNHp complex is shown with the N-terminal arch of Kap- β 2 lying on the horizontal plane, with an oblique view into the opening of the arch. The colour scheme is similar to Fig. 4a, with the addition of residues on Ran interacting with RBD1 (ref. 23) coloured purple and the Ran C-terminal extension in cyan. Side chains of N154 and E158 in Ran are purple while their main chains are green, as these residues interact with both Kap- β 2 and RBD1.

the intermolecular interactions are listed in Fig. 4b. Kap- β 2 residues at this interface are located within the first 170 residues of the protein, confirming previous deletion mutant studies that identified the N-terminal region of Kap- β 1 as an important interaction site for Ran-GTP^{26–28}. The C-terminal part of switch 1 (residues 44–47) of Ran interacts with Kap- β 2 helix N α B and the rest of this loop is exposed to solvent (Fig. 5). In contrast, most of switch 2 (residues 72–82) is buried at the interface through a series of hydrophobic contacts to HR1 α B of the karyopherin. Both polar and hydrophobic interactions are found at the N interface and most of the polar contacts are between arginine side chains on helix α 3 of Ran (R106 and R110) and serine and acidic residues on Kap- β 2 HR3 α B (S165, D164 and E161). Weak electron density (see Methods) and high B -factors in the N-terminal region of Kap- β 2 (B_{average} of 117 Å², residues 3–190; compared to B_{average} of 56 Å², residues 412–890), as well as in Ran switch 2 (B_{average} of 103 Å² for residues 69–83, compared to B_{average} of 74 Å² for entire Ran) indicate flexibility throughout much of the N interface. The preference of the N interface for Ran-GTP is obvious, as strand β 2E in Ran-GDP would clash sterically with N α B and prevent interactions between switch 1 and Kap- β 2 (Figs 3 and 5).

In contrast to the N interface, the central interface is characterized by strong electron density and interactions between well-ordered residues in both proteins, despite the lack of regular secondary structure in most of L7 (Fig. 3c). Figure 4a shows the location of residues at the interface and the interactions are listed in Fig. 4b. The central interface has a buried surface area of 2,400 Å² (62% of the total interface). Contacts in this region involve residues in loop L7 and the α B helices of HR6–HR7 and HR13–HR14 of the karyopherin with the face of Ran composed of λ 8, α 4, λ 9, β 6, λ 10 and λ 11, consistent with mapping of the Ran-binding domain to the N-terminal half of Kap- β 1 (refs 26–30). The importance of L7 in Ran binding was previously indicated when removal of residues 329–342 of Kap- β 1 (equivalent to Kap- β 2 361–374) significantly decreased Ran binding²⁹. Many of the central interface residues on Ran are buried in Ran-GDP by the C-terminal extension packing against the G-domain (Figs 3, 4). In the Kap- β 2 complex, the

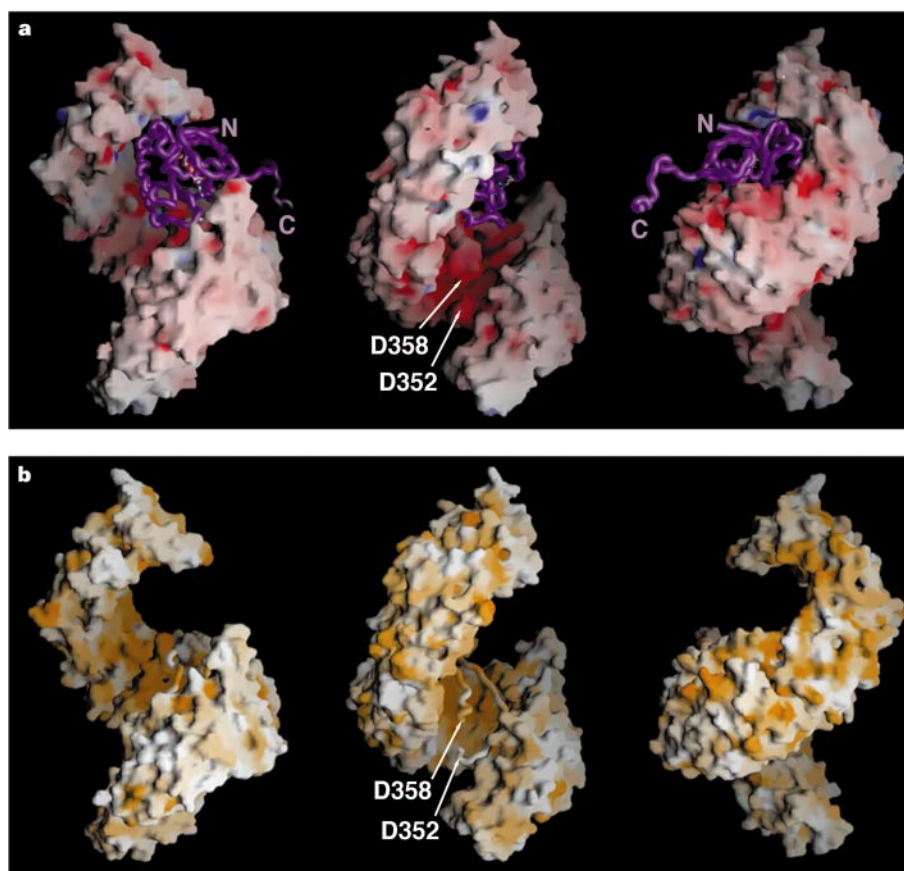


Figure 6 Electrostatic surface potential and sequence conservation mapped onto the molecular surface of Kap-β2 in the Kap-β2-Ran-GppNHp complex. **a**, Electrostatic surface potential of Kap-β2. Negative potential is red, uncharged surface is white and positive potential is blue. The left-hand view is oriented as in Figs 1, 2a and 4a. Additional views are generated by two successive 90° rotations about the vertical axis. Ran-GppNHp is represented by its backbone tube in purple. **b**, Sequence identity between Kap-β2s from seven different organisms

(human, *Xenopus laevis*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*). Sequence identity ≤30% is shown as white and sequence identity of 30–100% is represented with a colour ramp of white to gold. The orientation and views are as in **a**. Ran is removed and residues 333–372 of L7 on Kap-β2 are shown as backbone tube, with sequence identity mapped onto the tube. This figure was generated using GRASP⁵⁰.

C-terminal extension of Ran is displaced and the underlying surface becomes available for Kap-β2 binding. Thus, the nucleotide-switch-induced flexibility of the C-terminal extension of Ran-GTP not only allows that region of the GTPase to bind effector RBD1 (ref. 23) but also facilitates exposure of previously inaccessible interaction sites for Kap-β effectors.

The α4–β6 region of Ran is buried at the Kap-β2 central interface and shows structural differences (residues 137–142) when compared to Ran-GDP¹⁶ (Figs 3, 4a). The short α4 helix in Ran-GDP is further distorted in the complex to place H139 and R140 in more exposed conformations for intermolecular interactions. This region of the GTPase contains a basic patch (residues 139–142) that has been suggested to interact with the DEDDDL acidic patch at the C terminus of Ran-GDP²³. However, in the Kap-β2 complex, this basic patch on Ran does not contact similar acidic patches in L7 (³²¹DVEEDE³²⁶ and ³⁵⁰DEDGIEEDDDDDDEIDDD³⁶⁸). Instead, Ran H139 interacts with I370 in L7, and R140 in Ran makes numerous contacts to helices HR6αB (E278) and HR7αB (K377 and S378) in the karyopherin (Fig. 4b). The structural changes within the α4–β6 region of Ran may be induced by complexing with Kap-β2, or they may reflect inherent structural plasticity in the GTPase, as residues 137–143 vary between different Ran-GDP molecules in an asymmetric unit and between different Ran-GDP structures¹⁶.

Substrate-binding site and dissociation

The binding site on Kap-β2 for the A1-NLS has been mapped in

different studies to residues 518–890 (ref. 13) and 541–890 (ref. 14). These residues are located in the C-terminal arch of Kap-β2, spanning HEAT repeat HR10 to the C terminus (Fig. 2). To locate the binding site more precisely, we calculated the electrostatic surface potential and analysed the sequence conservation of Kap-β2 (Fig. 6).

At physiological pH, Kap-β2 is negatively charged (isoelectric point 4.7). The electrostatic potential surface of the protein shows that most of the negative charge is concentrated on the concave surface of the C-terminal arch, contributed by αB helices of HEAT repeats HR7–HR17 (Fig. 6a). The A1-NLS is rich in glycines and asparagines, and contains several basic residues, giving it an isoelectric point of 10.7. Complementarity between the positively charged NLS and the negatively charged interior of the Kap-β2 C-terminal arch suggests that A1 interacts with the concave surface of the C-terminal arch. We have also aligned Kap-β2s from seven different organisms and mapped the conservation of each residue onto the structure (Fig. 6b). A continuous band of high sequence identity is observed throughout the concave surface of the Kap-β2 spiral in both the N- and C-terminal arches. The asymmetric distribution of sequence conservation is consistent with the location of function conserved in Kap-β2 proteins, with Ran binding in the N-terminal arch and substrate binding in the C-terminal arch.

Biochemical studies have demonstrated that binding of Ran-GTP and the import substrate to Kap-β2 are mutually exclusive^{9,10}, leading to a model in which the GTPase triggers substrate release into the nucleus. Thus, physical coupling of the Ran and substrate

binding sites is critical to Kap- β 2 function. Identification of these sites in the structure indicates a molecular mechanism by which this could occur. Coupling appears to be mediated by the large intra-HEAT loop L7, which spans both ligand-binding sites. Two segments of L7 contact the main body of Kap- β 2: residues 320–330 pack across the N termini of the α B helices of HR9–HR11, and residues 340–350 pack against the α B helices of HR12–HR17 (Figs 2a, 4a). The latter segment covers a large portion of the C-terminal arch, including the putative substrate-binding site. In contrast, residues between (333–340) and C-terminal to (363–373) these segments do not contact Kap- β 2, but interact extensively with loops λ 8 and λ 10 (residues 124–127, 153–159), and residues spanning helix α 4–strand β 6 (129–146) of Ran, respectively (Fig. 4). Thus, it is likely that Ran binding has a strong influence on the structure and orientation of L7, and therefore on the interactions of the substrate with the C-terminal arch. In the absence of a Kap- β 2–substrate complex structure, we cannot assess the detailed nature of karyopherin–substrate interactions and Ran–substrate coupling. One possibility is that in the absence of Ran, loop L7 is disordered, and substrate binds directly to helices lining the C-terminal arch. Ran binding could then pack the loop against HEAT repeats HR12–HR17, displacing the substrate. Alternatively, the loop itself, which is very acidic (Fig. 2b) and could thus complement the basic A1–NLS sequence, may be part of the substrate-binding site. In this case, Ran binding could simply alter the loop conformation to displace the substrate. This possibility is supported by the observation that deletion of residues 329–342 of Kap- β 1 (361–374 in Kap- β 2) decreases affinity for Kap- α significantly²⁹. Structures of free Kap- β 2 and the Kap- β 2–substrate complex will allow us to confirm the roles of loop L7 in substrate binding and GTPase-mediated substrate dissociation.

Higher order Kap- β 2–Ran complex interactions

When complexed with Kap- β s, Ran-GTP is resistant to both ranGAP-mediated GTPase activation and ranGEF-mediated nucleotide exchange^{31–33}. The structures of rasGAP and rhoGAP complexed with their respective GTPases^{34,35} reveal similar modes of GAP–GTPase interaction and GTPase-activation mechanisms, despite the absence of similarity in primary or tertiary structures between the GAPs. In both cases, GTPases interact with GAPs primarily through switch 1 and switch 2. As Ran-GppNHp in the Kap- β 2 complex resembles Ras-GTP, and GAP–GTPase interactions seem conserved across different GTPases, it is likely that Ran will also interact with ranGAP through its switch regions. However, as residues 43–47 in switch 1 and most of switch 2 in Ran are buried at the N interface of the complex (Fig. 4a), these regions are inaccessible for predicted interactions with ranGAP. Ras also interacts extensively with its GEF, Sos, through switch 1 and switch 2 (ref. 36). Assuming similar interactions in ranGEF-mediated nucleotide exchange, Kap- β 2–Ran complex formation would also prevent interactions between the switch regions of Ran and ranGEF, thereby inhibiting ranGEF-induced nucleotide exchange.

As well as Kap- β s, Ran-GTP also binds ranBP1 (refs 28, 33, 37) and functionally similar RBH domains such as RBD1 (ref. 23) in Nup358 (refs 28, 29). ranBP1 is required for nuclear transport but its role remains uncertain. The RBH domains of Nup358 may target Ran-GTP to the NPC for hydrolysis by ranGAP, and Kap- β –Ran–RBH complexes may concentrate transport factors to the cytoplasmic fibres of the NPC in preparation for import¹. The recently reported structure of the RBD1–Ran-GppNHp complex revealed that the GTPase interacts with this target through residues in switch 1 (29–38), N55 in the β 2– β 3 loop, N154 and E158 in λ 10 and residues in the C-terminal extension²³ (Fig. 5). Of these, only N154 and E158 interact with Kap- β 2 in our structure. Residues 154 and 158 flank one edge of the Kap- β 2 central interface and their side chains are directed away from the interface (Fig. 5). In the RBD1 complex, these residues contact effector residues R82, E83 and Q84,

which are part of a loop that extends from the core of RBD1 (ref. 23). Therefore, it appears that the Kap- β 2 and RBD1 interfaces are mostly non-overlapping (Fig. 5). This is consistent with biochemically observed ternary complexes involving Ran-GTP, karyopherin and ranBP1 (refs 28, 33). However, the two effector interfaces on Ran are adjacent and, in the ternary complex, competition between the effectors for interactions at Ran residues N154 and E158 may attenuate Kap- β 2–Ran interactions⁴⁰ (Fig. 5).

ranBP1 binds Ran-GTP with nanomolar K_D , and its affinity for Ran-GDP is approximately 10 μ M (ref. 41). Nevertheless, Kap- β 1 can form stable ternary complexes with both Ran-GTP–ranBP1 and Ran-GDP–ranBP1 (refs 28, 33). Similar ternary complexes can be formed with Kap- β 2 (data not shown). The Kap- β 2–Ran structure, in combination with biochemical studies involving ranBP1 (ref. 41) and the structure of the Ran-GppNHp–RBD1 complex²³, offers a rationale for the formation of a Kap- β 2–Ran-GDP–ranBP1 ternary complex, despite the low affinities of Ran-GDP for either Kap- β 2 or ranBP1 alone. The C-terminal extension of Ran interacts extensively with RBD1 (ref. 23). In the ranBP1–Ran-GDP complex, the energetic penalty for dissociating the C terminus from Ran could contribute to the decrease in affinity. Our structure shows displacement of the Ran C terminus and extensive contacts between Kap- β 2 and regions of Ran lying under the C-terminal extension in the GDP state (Figs 3, 4a, b). Therefore, in the Kap- β 2–Ran-GDP–ranBP1 ternary complex, displacement of the C terminus by either Kap- β 2 or ranBP1 would facilitate interactions of the other protein, thereby stabilizing the trimeric structure.

Discussion

The helical HEAT repeats in Kap- β 2 assemble to provide a large surface for interactions with multiple ligands; these relatively simple structural motifs can be organized into orthogonal and spatially distinct binding sites for GTPase and substrate. The 62-residue intra-HEAT repeat loop L7 interacts extensively with Ran through some of its residues, and adjacent portions of the loop interact with the substrate-binding site. This arrangement suggests an allosteric mechanism whereby the loop could transmit structural information between the two ligand-binding sites. Occupation of the loop in the C-terminal arch may displace substrate from a similar or overlapping site, allowing it to be released into the nucleus.

Members of the karyopherin- β family have similar molecular weights and isoelectric points, and contain multiple HEAT repeats^{1,21}. The structure of Kap- β 2 is therefore a likely archetype for this family of transport proteins. Kap- β s are most conserved at their N termini¹, indicating that the mode of GTPase binding observed here may be similar throughout the family. However, these proteins bind diverse classes of transport substrates with the C-terminal half of their sequences^{13,14,26} and it will be important to understand how their substrate-binding sites are adapted to recognize such different ligands as proteins and RNA¹. Sequence analysis of Kap- β 1 and Kap- β 3 also shows a largely hydrophilic peptide stretch of about 60 residues (270–343, Kap- β 1; 266–344, Kap- β 3) which aligns with the Kap- β 2 L7 loop. This indicates that similarities in the Kap- β homologues may include the use of a long polypeptide in extended conformation to mediate GTPase signals for both nuclear import and nuclear export. □

Methods

Expression and protein purification. The human Kap- β 2 gene (residues 1–890)¹² was subcloned into pGEX4T3 (Pharmacia) containing a previously added Tev protease-cleavage site and expressed in *Escherichia coli* BL21-DE3. Bacteria were grown as described to produce selenomethionine-labelled protein⁴². Cells resuspended in buffer A (20 mM HEPES, pH 7.3, 110 mM potassium acetate, 1 mM EDTA, 10 mM DTT, 20% glycerol) were lysed in a French pressure cell. The lysate was loaded onto glutathione-Sepharose (Pharmacia) and GST–Kap- β 2 was eluted with buffer A containing 20 mM glutathione, pH 8.2. Concentrated GST–Kap- β 2 was cleaved with Tev protease

(GibcoBRL) and then purified on Q-HiTrap and Superdex-200 columns (Pharmacia). *E. coli* BL21-DE3 transformed with pET11d-human Ran (residues 1–216) were grown as above. We purified Ran from a French press lysate in buffer B (20 mM HEPES, pH 7.3, 110 mM potassium acetate, 5 mM magnesium acetate, 1 mM EGTA, 10 mM DTT, 20% glycerol) containing 10 μ M GDP, on Q-Hi-Trap, SP Hi-Trap and Superdex-75 columns (Pharmacia). Selenomethionine Kap- β 2–Ran-GppNHP complex was formed after incubating selenomethionine Kap- β 2 with 2–3-fold excess selenomethionine Ran-GppNHP (obtained by nucleotide exchange). The selenomethionine Kap- β 2–Ran-GppNHP complex was purified by gel filtration on Superdex-200 (Pharmacia).

Crystallization and data collection. Crystals were grown in hanging drops with reservoir solution containing 100 mM HEPES, pH 7.5, 28% PEG3350 and 0.2 M $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$. Showers of rod-shape crystals appear after 24 h. Most remain at initial dimensions of 20 $\mu\text{m} \times 20 \mu\text{m} \times 1 \text{mm}$, but a few grew in two weeks to 60 $\mu\text{m} \times 60 \mu\text{m} \times 1 \text{mm}$. The larger crystals were harvested in a 1:1 mixture of buffer B and reservoir solution, cryoprotected with 17% PEG3350 and 15% glycerol and flash-frozen in liquid propane. All datasets were collected on Beamline X25 at Brookhaven National Laboratory using the Brandeis 2×2 CCD-based detector⁴³ and processed with DENZO/SCALEPACK⁴⁴. Seven selenium sites were located with the SHELXS-97 Patterson interpretation program¹⁸ using dispersive and anomalous differences. Heavy-atom positions were refined and phases calculated with MLPHARE⁴⁵. Difference Fourier methods located 17 more selenium sites (total 24 of 30 possible sites). Density modification with DM and SOLOMON⁴⁵ resulted in interpretable electron-density maps.

Model building and refinement. The model was built using program O⁴⁶. The initial model included residues 170–352 and 368–889 of Kap- β 2, residues 8–73, 79–177 and 181–190 of Ran, GppNHP, Mg^{2+} and polyaniline traces of helices (without connecting loops) spanning residues 3–166 of Kap- β 2. Despite weak density at the N-terminal region of Kap- β 2 (residues 3–166), these densities show unambiguously the α -helices in that region. The Ran model was built with the guidance of a published Ran-GDP structure¹⁶. Refinement was performed with CNS using conjugate gradient minimization, torsion angle simulated annealing and restrained individual *B*-factor refinement, using all data from 500–3.0 Å (ref. 47). MAD phases were included in the refinement using the MLHL maximum likelihood target, and refinement was performed with bulk solvent correction and anisotropic *B*-factor corrections (final tensor elements of $B_{11} = -9.56 \text{ Å}^2$, $B_{33} = 19.119 \text{ Å}^2$, $B_{12} = -13.223 \text{ Å}^2$). Phase combination of the refined model with experimental MAD phases using Sigmaa⁴⁵ allowed most of the remaining regions of Kap- β 2 and Ran to be built. Electron density at Kap- β 2 residues 3–166 and Ran residues 74–78 is continuous but still weak, with obvious side-chain densities only when contoured at 0.6–0.7 σ , indicating significant flexibility and mobility in those regions. Residues 1–2, 167–169 and 153–157 of Kap- β 2, and residues 1–7 and 198–216 of Ran are disordered and not included in the model. The current *R*-value is 26.7% (*R*_{free} of 31.5%; 500.0 and 3.0 Å, no amplitude cutoff); the structure shows good stereochemistry and no Ramachandran violations are observed.

Received 11 March; accepted 26 April 1999.

- Pemberton, L. F., Blobel, G. & Rosenblum, J. S. Transport routes through the nuclear pore complex. *Curr. Opin. Cell Biol.* **10**, 392–399 (1998).
- Nehrbass, U. & Blobel, G. Role of the nuclear transport factor p10 in nuclear import. *Science* **272**, 120–122 (1996).
- Ribbeck, K., Lipowsky, G., Kent, H. M., Stewart, M. & Gorlich, D. NTF2 mediates nuclear import of Ran. *EMBO J.* **17**, 6587–6598 (1998).
- Conti, E., Uy, M., Leighton, L., Blobel, G. & Kuriyan, J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin α . *Cell* **94**, 193–204 (1998).
- Albertini, M., Pemberton, L. F., Rosenblum, J. S. & Blobel, G. A novel nuclear import pathway for the transcription factor TFIIS. *J. Cell Biol.* **143**, 1447–1455 (1998).
- Kaffman, A., Rank, N. M. & O'Shea, E. K. Phosphorylation regulates association of the transcription factor Ph04 with its import receptor Pso1/Kap121. *Genes Dev.* **12**, 2673–2683 (1998).
- Kaffman, A., Rank, N. M., O'Neill, E. M., Huang, L. S. & O'Shea, E. K. The receptor MsN5 exports the phosphorylated transcription factor Ph04 out of the nucleus. *Nature* **396**, 482–486 (1998).
- Rexach, M. & Blobel, G. Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell* **83**, 683–692 (1995).
- Siomi, M. C. *et al.* Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins. *J. Cell Biol.* **138**, 1181–1192 (1997).
- Izaurrealde, E., Kutay, U., von Kobbe, C., Mattaj, J. W. & Gorlich, D. The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *EMBO J.* **16**, 6535–6547 (1997).
- Bischoff, F. R. & Ponstingl, H. Catalysis of guanine nucleotide exchange of Ran by RCC1 and stimulation of hydrolysis of Ran-bound GTP by Ran-GAP1. *Methods Enzymol.* **257**, 135–144 (1995).
- Bonifaci, N., Morioianu, J., Radu, A. & Blobel, G. Karyopherin β 2 mediates nuclear import of a mRNA binding protein. *Proc. Natl Acad. Sci. USA* **94**, 5055–5060 (1997).

- Pollard, V. W. *et al.* A novel receptor-mediated nuclear protein import pathway. *Cell* **86**, 985–994 (1996).
- Fridell, R. A., Truant, R., Thorne, L., Benson, R. E. & Cullen, B. R. Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin- β . *J. Cell Sci.* **110**, 1325–1331 (1997).
- Andrade, M. A. & Bork, P. HEAT repeats in Huntington's disease protein. *Nature Genet.* **11**, 115–116 (1995).
- Stewart, M., Kent, H. M. & McCoy, A. J. The structure of the Q69L mutant of GDP-Ran shows a major conformational change in the switchII loop that accounts for its failure to bind nuclear transport factor2 (NTF2). *J. Mol. Biol.* **284**, 1517–1527 (1998).
- Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. & Wittinghofer, A. Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. *Nature* **374**, 378–381 (1995).
- Sheldrick, G. M., Dauter, Z., Wilson, K. S., Hope, H. & Sieker, L. C. The application of direct methods and Patterson interpretation to high-resolution native protein data. *Acta Crystallogr. D* **49**, 18–23 (1993).
- Groves, M. R., Hanlon, N., Turowski, P., Hemmings, B. A. & Barford, D. The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 randomly repeated HEAT motifs. *Cell* **96**, 99–110 (1999).
- Huber, A. H., Nelson, W. J. & Weiss, W. I. Three-dimensional structure of the armadillo repeat region of β -catenin. *Cell* **90**, 871–882 (1997).
- Malik, H. S., Eickbush, T. H. & Goldfarb, D. S. Evolutionary specialization of the nuclear targeting apparatus. *Proc. Natl Acad. Sci. USA* **94**, 13738–3742 (1997).
- Richards, S. A., Lounsbury, K. M. & Macara, I. G. The C terminus of the nuclear RAN/TC4 GTPase stabilizes the GDP-bound state and mediates interactions with RCC1, RAN-GAP, and HTF9A/RANBP1. *J. Biol. Chem.* **270**, 14405–14411 (1995).
- Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39–46 (1999).
- Pai, E. F. *et al.* Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. *Nature* **341**, 209–214 (1989).
- Hidea, M. *et al.* A monoclonal antibody to the COOH-terminal acidic portion of Ran inhibits both the recycling of Ran and nuclear protein import in living cells. *J. Cell Biol.* **144**, 645–655 (1999).
- Kutay, U., Izaurrealde, E., Bischoff, F. R., Mattaj, J. W. & Gorlich, D. Dominant-negative mutants of import- β block multiple pathways of import and export through the nuclear pore complex. *EMBO J.* **16**, 1153–1163 (1997).
- Kose, S., Imamoto, N., Tachibana, T., Shimamoto, T. & Yoneda, Y. Ran-unassisted nuclear migration of a 97-kD component of nuclear pore-targeting complex. *J. Cell Biol.* **139**, 841–849 (1997).
- Chi, N. C., Adam, E. J. H. & Adam, S. A. Different binding domains for Ran-GTP and Ran-GDP/RanBP1 on nuclear import factor p97. *J. Biol. Chem.* **272**, 6818–6822 (1997).
- Morioianu, J., Blobel, G. & Radu, A. Nuclear protein import: Ran-GTP dissociates the karyopherin α heterodimer by displacing α from an overlapping binding site on β . *Proc. Natl Acad. Sci. USA* **93**, 7059–7062 (1996).
- Enkel, C., Schulke, N. & Blobel, G. Expression in yeast of binding regions of karyopherins α and β and inhibits nuclear import and cell growth. *Proc. Natl Acad. Sci. USA* **93**, 12986–12991 (1996).
- Gorlich, D., Pante, N., Kutay, U., Aebi, U. & Bischoff, F. R. Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J.* **15**, 5584–5594 (1996).
- Floor, M. & Blobel, G. The nuclear transport factor karyopherin β binds stoichiometrically to Ran-GTP and inhibits the Ran GTPase activating protein. *J. Biol. Chem.* **271**, 5313–5316 (1996).
- Lounsbury, K. M. & Macara, I. G. Ran-binding protein 1 (RanBP1) forms a ternary complex with Ran and karyopherin β and reduces Ran GTPase-activating protein (RanGAP) inhibition by karyopherin β . *J. Biol. Chem.* **272**, 551–555 (1997).
- Scheffzek, K. *et al.* The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* **277**, 333–338 (1997).
- Rittinger, K., Walker, P. A., Eccleston, J. F., Smerdon, S. J. & Gamblin, S. J. Structure at 1.65 Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue. *Nature* **389**, 758–762 (1997).
- Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, K. The structural basis of the activation of Ras by Sos. *Nature* **394**, 337–343 (1998).
- Coutavas, E., Ren, M., Oppenheim, J. D., Eustachio, P. & Rush, M. G. Characterization of proteins that interact with the cell-cycle regulatory protein Ran/TC4. *Nature* **366**, 585–587 (1993).
- Wu, J., Matunis, M. J., Kraemer, D., Blobel, G. & Coutavas, E. Nup358, a cytoplasmically exposed nucleoporin with peptide repeats, Ran-GTP binding sites, zinc fingers, a cyclophilin A homologous domain, and a leucine-rich region. *J. Biol. Chem.* **270**, 14209–14213 (1995).
- Yokoyama, N. *et al.* A giant nucleopore protein that binds Ran/TC4. *Nature* **376**, 184–188 (1995).
- Bischoff, F. R. & Gorlich, D. RanBP1 is crucial for the release of RanGTP from importin beta-related nuclear transport factors. *FEBS Lett.* **419**, 249–254 (1997).
- Kuhlmann, J., Macara, I. & Wittinghofer, A. Dynamic and equilibrium studies on the interaction of Ran with its effector, RanBP1. *Biochemistry* **36**, 12027–12035 (1997).
- Van Duyn, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structure of FKBP-FK506, an immunophilin-immunosuppressant complex. *Science* **251**, 839–842 (1991).
- Westbrook, E. M. & Naday, I. Charge-coupled device-based area detectors. *Methods Enzymol.* **276**, 244–268 (1997).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation model. *Methods Enzymol.* **276**, 307–326 (1997).
- CCP4. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
- Jones, T. A., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Kraulis, P. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merritt, E. A. & Murphy, M. E. P. Raster3D Version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Acknowledgements. We thank P. Jeffreys and N. Pavletich of MSKCC and J. Kuriyan of RU for the use of X-ray data collection facilities; R. Sweet and L. Berman of NSLS and M. Rosen for help with synchrotron data collection; J. Kuriyan and T. Sakmar for use of computation facility; N. Bonifaci for the clone of Kap β 2; and M. Rosen, E. Conti, J. Bonanno, P. Jeffreys, P. Adams, K. Reinisch, J. Helmers, L. Pemberton, J. Rosenblum, M. Floor, M. Rout and J. Goldberg for advice and help. Y.M.C. is supported by a Life Sciences Research Foundation Fellowship sponsored by the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to Y.M.C. The atomic coordinates have been deposited in the Brookhaven Protein Data Bank, accession number 1QBK.

Observation of quasiparticles with one-fifth of an electron's charge

M. Reznikov^{*†}, R. de Picciotto^{*†}, T. G. Griffiths^{*}, M. Heiblum^{*} & V. Umansky^{*}

^{*} Braun Center for Submicron Research, Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 76100, Israel

The fractional quantum Hall effect¹ occurs in the conduction properties of a two-dimensional electron gas subjected to a strong perpendicular magnetic field. In this regime, the Hall conductance shows plateaux, or fractional states, at rational fractional multiples of e^2/h , where e is the charge of an electron and h is Planck's constant. The explanation^{1–3} of this behaviour invokes strong Coulomb interactions among the electrons that give rise to fractionally charged quasiparticles which can be regarded as non-interacting current carriers^{1–5}. Previous studies^{4,5} have demonstrated the existence of quasiparticles with one-third of an electron's charge, the same fraction as that of the respective fractional state. An outstanding ambiguity is therefore whether these studies measured the charge or the conductance. Here we report the observation of quasiparticles with a charge of $e/5$ in the $2/5$ fractional state, from measurements of shot noise in a two-dimensional electron gas⁴. Our results imply that charge can be measured independently of conductance in the fractional quantum Hall regime, generalizing previous observations of fractionally charged quasiparticles.

In the fractional quantum Hall (FQH) regime, the first Landau level is partly populated, or 'fractionally filled'. Laughlin's explanation of the FQH effect^{1–3} involved the emergence of new, fractionally charged, quasiparticles. Shot-noise measurements^{4,5} have confirmed the existence of these quasiparticles in the FQH regime. Shot noise, resulting from the granular nature of the particles, is proportional to the charge of the current carriers, in this case quasiparticles^{4,5}. In these experiments a quantum point contact (QPC) embedded in a two-dimensional electron gas (2DEG), serving as a potential constriction, was used as an electronic 'beam splitter'. Its purpose is to partly reflect back the incoming current and lead to partitioning of carriers and hence to shot noise. An applied magnetic field corresponding to fractional filling factors (in the bulk far from the QPC) of $\nu_B = 1/3$ in ref. 4 or $2/3$ in ref. 5, was employed. Charge was deduced via the generalized equation for shot noise of non-interacting particles (the classical and simplified version is the Schottky formula: $S = 2qI_B$, with S the low-frequency spectral density of current fluctuations, I_B the reflected current, and q the charge of the current-carrying particle). For small reflection by the QPC (small I_B) the quasiparticle's charge was found to be $e/3$ (ref. 4 and 5), as predicted theoretically^{6–8}. The theories are based on the chiral Luttinger-liquid model and are applicable only for Laughlin fractional states, for example, $1/3$, $1/5$, and so on. For other, more general filling factors (such as $\nu = 2/5$), however, such calculations become exceedingly complicated. Still, one can gain insight into the characteristics of the expected shot noise in such cases by considering the more intuitive composite fermion (CF) model^{9,10}.

Laughlin suggested that in the FQH regime the current is carried by weakly interacting quasiparticles with charge $q = e/(2pn + 1)$, where e is the electronic charge. The fractional filling factor $\nu = p/(2pn + 1)$ determines the conductance of the sample $g = \nu g_0$, with $g_0 = e^2/h$ being the quantum conductance. Within

the CF model, fractional filling factors for electrons of the form $\nu = p/(2pn + 1)$ are identified as integer filling factors of CFs, $\nu_{CF} = p$. Each CF is composed of a single electron with $2n$ quanta of magnetic flux (derived from the applied magnetic field) attached to it (each flux quantum is $\phi_0 = h/e$). Here we deal with the simplest family of CFs, with only two flux quanta attached to each electron ($n = 1$). A filling factor $\nu = 1/3$ then corresponds to the simplest fraction $p = 1$ while $\nu = 2/5$ corresponds to $p = 2$, that is, two CF Landau levels are filled. The effective magnetic field sensed by the CFs is $B - 2n_s(h/e)$, with n_s the density of the 2DEG and B the magnetic field. Under this, weaker, effective magnetic field the CFs are usually considered as non-interacting quasiparticles (though a justification of this assertion is not solid). Shot noise, induced by the QPC, is thus recognized as partition noise—associated with partial reflection of CFs in integer Landau levels (referred to below as 'CF channels'). This scheme is similar to that taking place at a QPC reflector at zero applied magnetic field^{11–14}. The shot noise in a weakly reflected channel p (filling factor $p/(2p + 1)$) is expected¹⁵ to correspond to quasiparticles with charge $q = e/(2p + 1)$.

It is important to examine the noise properties at fractional filling factors in which the value of the charge in units of e is expected to differ from the filling factor. Observation of a charge $e/(2p + 1)$ for $p > 1$ would prove that our charge measurement is not simply a different way to measure the conductance (or the filling factor)¹⁶. An obvious experimental target is the filling factor $\nu = 2/5$, where the charge of the current-carrying quasiparticles is expected to be $(1/5)e$ while the conductance is $(2/5)g_0$.

The contribution of each one-dimensional channel without magnetic field is known to be equal to the quantum conductance $g_0 = e^2/h$. However, the conductance of each CF channel has a

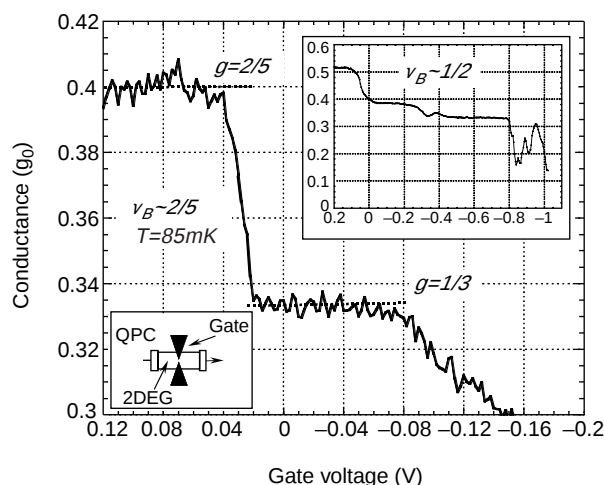


Figure 1 Two-terminal conductance (in units of the quantum conductance) versus voltage applied to the gates of the quantum point contact (QPC). Clear plateaux of $(2/5)g_0$ and $(1/3)g_0$ are seen for bulk filling factor $\nu_B = 2/5$ ($B = 12$ T). Top inset, a similar two-terminal conductance plot for a bulk filling factor of $\nu_B = 1/2$ ($B = 10.5$ T). Note that for zero voltage on the gates of the QPC the value of the conductance, in both cases, was lower than the bulk value ($g < \nu g_0$). This is a result of unintentional reflection from the QPC even when it is not biased. Indeed, when a small positive gate voltage was applied to the QPC the conductance increased and reached its bulk value. Application of a negative gate voltage recovered the $2/5$ and $1/3$ states, as expected. Bottom inset, schematic of the QPC in 2DEG. The QPC was formed by evaporating two metallic gates onto the surface of the heterostructure. By applying bias to the gates with respect to the 2DEG, the transmission of the incoming current through the small opening of the QPC is controlled. The electron temperature was determined by comparing the thermal noise without an external current through the QPC to the Johnson-Nyquist formula $S = 4k_B T g$ via $T = (\delta S / \delta g) / 4k_B$, with g the total conductance. Moreover, extrapolating this linear dependence to zero conductance (of the QPC) gives the contribution of our preamplifier to the total noise.

[†] Present addresses: Physics Department, Technion, Haifa 32000, Israel (M.R.); Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA (R.d.P.).

smaller, channel-dependent, value δg_p . For example, the contribution of the first CF channel to the conductance is $\delta g_1 = g_0/3$ while that of the second channel is $\delta g_2 = (\frac{2}{5} - \frac{1}{3})g_0$. Hence, when the two lowest CF channels are fully transmitted the higher channel carries only a small portion (1/6) of the total current. Taking into account the smaller charge of the corresponding quasiparticle, $e/5$, a very weak shot-noise signal is expected. This makes the measurement of the noise in the 2/5 channel exceedingly difficult.

Our experiment was performed with a set-up similar to the one used in refs 4 and 17. Noise was measured within a bandwidth of ~ 30 kHz about a central frequency of 1.68 MHz. This frequency was chosen by tuning a LRC resonant circuit, with C the capacitance of the coaxial cables and R being mainly the resistance of the QPC. By choosing a two-fold lower frequency compared to the one used previously (4 MHz)⁴ we reduced the spurious noise contribution of our preamplifier from $1.1 \times 10^{-28} \text{ A}^2 \text{ Hz}^{-1}$ before to $3.8 \times 10^{-29} \text{ A}^2 \text{ Hz}^{-1}$. The preamplifier, which operates at 4.2 K, was manufactured from transistors fabricated with a GaAs-AlGaAs heterostructure grown in our own molecular beam epitaxy system. Measurements were performed in a dilution refrigerator at an electron temperature $T = 85 \text{ mK}$.

The 2DEG was embedded in a GaAs-AlGaAs heterostructure with a low-temperature carrier concentration $n_s = 1.15 \times 10^{11} \text{ cm}^{-2}$ and mobility $4.2 \times 10^6 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. A perpendicular magnetic field of $\sim 12 \text{ T}$ led to a bulk filling factor 2/5. The measured two-terminal conductance of two different runs, at bulk filling factors $\nu = 2/5$ and $\nu \approx 1/2$, as a function of applied voltage to the QPC gates, are shown in Fig. 1. Two clear conductance plateaux are seen near $g = (2/5)g_0$ and $g = (1/3)g_0$.

In the absence of an exact model for shot noise at $\nu = 2/5$, we compare our results with the partition noise expected from non-interacting particles. In other words, we assume that the scattering events of the quasiparticles at the QPC are independent. This assumption has been quite successful in analysing the noise results for quasiparticles with charge $e/3$ in a single CF channel^{4,5}. At zero temperature ($T = 0$), one would expect the low-frequency spectral density of current fluctuations, S , to be given by:

$$S_{T=0} = 2qV\delta g_p t_p (1 - t_p) \quad (1)$$

with V the applied bias voltage across the QPC, δg_p the contribution of the p th CF channel to the total conductance, t_p the transmission

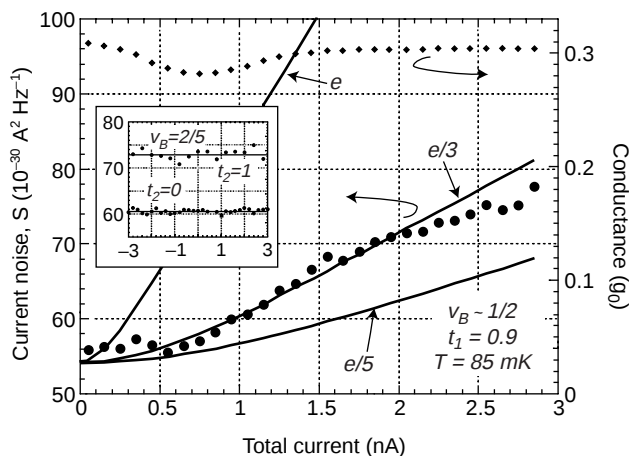


Figure 2 Measured spectral density of current fluctuations for $t_1 = 0.9, t_2 = 0$. The 2/5 channel is fully reflected (filled circles, left-hand vertical axis). The result agrees well with a quasiparticle charge $q = e/3$ and $T = 85 \text{ mK}$ substituted in equation (2) and is in excellent agreement with ref. 4. For comparison, the expected noise curves for quasiparticle charge $q = e/5$ and $q = e$ are shown. The inset shows the noise measured at both 2/5 ($t_1 = 1, t_2 = 1$) and 1/3 ($t_1 = 1, t_2 = 0$) plateaux. No excess noise is measured on the plateaux as no partitioning takes place and the reservoirs produce a noise-free current.

coefficient of this channel, and q the charge of the quasiparticle¹⁴. For example, for $p = 2$, $\delta g_2 = g - g_0/3$ and $t_2 = [(g/g_0) - 1/3]/(2/5 - 1/3)$.

A more subtle issue is the expected noise at a finite temperature T . This noise does not vanish at zero applied voltage but approaches the Johnson–Nyquist formula: $S = 4k_B T g$, in accordance with the fluctuation-dissipation theorem. When a bias voltage V is applied, the noise is expected to increase smoothly with increasing V , approaching the linear behaviour predicted by equation (1) at an applied voltage greater than $V_T \approx 2k_B T/q$. Analytical expressions which take heuristically into account fractional statistics of the quasiparticles but ignore interactions among them¹⁸ lead to results that are very close to the one derived for non-interacting fermions with the electronic charge e replaced by the quasiparticle charge. On the other hand, numerical calculations¹⁹ that were done only for the

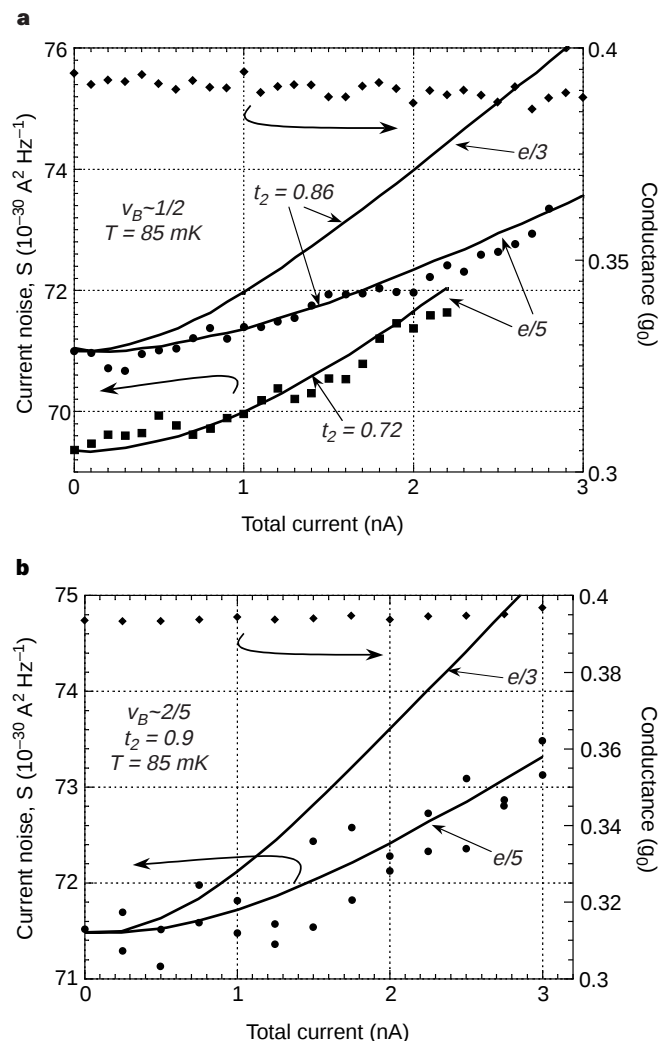


Figure 3 Measured noise of quasiparticles in the second composite fermion (CF) channel. The first CF channel (the '1/3' channel) is fully transmitted and does not produce noise (as seen Fig. 2 inset). **a**, Spectral density of current fluctuations against the total average current for transmission $t_2 = 0.86$ and $t_2 = 0.72$ at bulk filling factor $\nu_B = 1/2$ (filled circles, left-hand vertical axis). Solid lines are given by equation (2) assuming a charge $q = e/3$ and $q = e/5$ and $T = 85 \text{ mK}$ —as indicated. The sample's differential conductance, for transmission $t_2 = 0.86$ is also shown (filled diamonds, right-hand vertical axis). Note that the differential conductance, and hence the deduced transmission, are rather constant over the full range of the measurement. **b**, Similar noise and conductance data for $t_2 = 0.9$ and bulk filling factor $\nu_B = 2/5$. We note that there are no fitting parameters in the theoretical curves, as the transmission coefficient t_2 and the temperature of the electrons T are both measured independently^{4,17}.

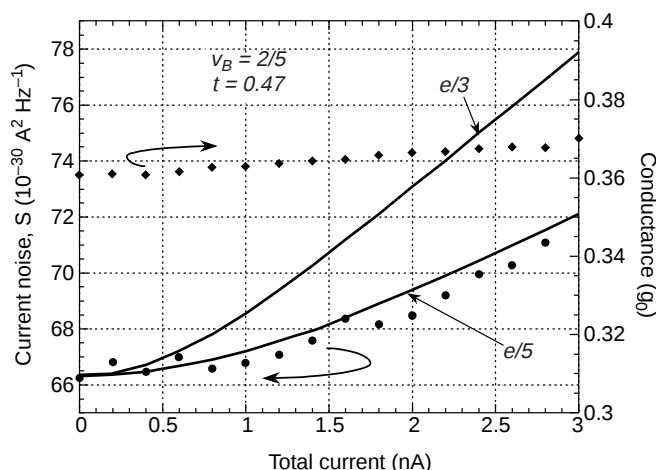


Figure 4 Current noise and conductance plotted against total average current. The measured spectral density of current fluctuations is given for relatively small transmission coefficient, $t_2 = 0.47$ (filled circles, left-hand vertical axis). Solid lines are given by equation (2) assuming charges of $q = e/3$ and $q = e/5$ —as indicated. The sample's differential conductance is also shown (filled diamonds, right-hand vertical axis). The transmission t_2 is determined by the average differential conductance over the full current range. We note, though, that the differential conductance in this case is not entirely independent of the applied voltage. In

general, a nonlinear I - V characteristic complicates the interpretation of our measurements: however, we expect the modifications to equation (2), due to the energy dependence of t_2 , are small. Apart from possible channel mixing, a change in the conductance by a relatively small amount induces a maximal change in the thermal noise $4k_B T \delta g$ of some 25% of the measured excess noise. Also, the weak dependence of the expected shot noise on t_2 near $t = 0.5$ (in a non-interacting-fermions picture) makes this nonlinearity insignificant.

fractional state $1/3$ contain parameters that are difficult to obtain experimentally. Hence we employ the expression for non-interacting particles²⁰ that was used successfully in the interpretation of the measurements in the $1/3$ fractional state⁴:

$$S = 2qV\delta g_p t_p (1 - t_p) \left[\tanh\left(\frac{qV}{2k_B T}\right) - \frac{2k_B T}{qV} \right] + 4k_B T g \quad (2)$$

We start with noise measurements on the conductance plateaux. The inset in Fig. 2 shows the results on both the $2/5$ ($t_2 = 1$, $t_1 = 1$) and $1/3$ ($t_2 = 0$, $t_1 = 1$) plateaux. We find no excess noise (above the thermal noise) within the accuracy of the measurement. This shows that the impinging current in both cases is noiseless, as one would expect for two independent CF channels. We come to this point later again. To tie the present results with previous experiments, we start with measurements of noise generated by partly reflecting the first CF channel (the $1/3$ channel). To validate our previous measurements of the charge $e/3$ and to verify that only the partitioning at the QPC is important in the determination of the charge, a different bulk filling factor, $\nu_B = 1/2$, was chosen (a bulk filling factor $1/3$ was employed before). Figure 2 shows the measured noise for a QPC partly reflecting the first CF channel and fully reflecting all the higher CF channels. The data, after calibration and subtraction of the amplifier's contribution, agree very well with the predicted shot noise with $q = e/3$.

The current noise for weakly back-scattered quasiparticles in the second CF channel, at two different bulk filling factors, is shown in Fig. 3. Figure 3a shows the measured noise for two values of the transmission of the QPC, $t_2 = 0.86$ and $t_2 = 0.72$, at a bulk filling factor $\nu_B \approx 1/2$. Using a quasiparticle charge $q = e/5$ and transmission deduced from the average value of the conductance (within the applied direct current range), the measured noise agrees well with the prediction of equation (2) throughout the whole range of direct current. Similarly, Fig. 3b shows the noise data for a different bulk filling, $\nu_B = 2/5$ and $t_2 = 0.9$. Even though the signal is weak and scattering of the data points is relatively large, it clearly verifies that the charge of the quasiparticles is $q = e/5$. Here, again, we see a clear manifestation of transport in the second CF channel with no contribution of the first CF channel (which is fully transmitted), as one would expect for non-interacting CF channels.

A similar agreement, even though surprising, between the measured noise and equation (2) is also obtained with a smaller transmission coefficient (within the range $0.4 < t_2 < 0.6$), as is demonstrated by the example in Fig. 4. Again a charge $q = e/5$ is measured. Theories^{6,7} find a smooth variation of the measured quasiparticle's charge from $q = e/3$ for weak back scattering to $q = e$ for strong back scattering, in the first CF channel ($\nu = 1/3 \rightarrow \nu = 0$ at the QPC). Extending such arguments to the transition $\nu = 2/5 \rightarrow \nu = 1/3$ within the QPC region, one would expect to determine via shot-noise measurements a charge $q = e/5$ at large t_2 changing monotonically to $q = e/3$ at small t_2 . Strong conductance variations with current for even stronger back reflection of the second CF channel ($t_2 < 0.4$) prevents accurate measurements of the shot noise and deduction of the value of the quasiparticle charge. □

Received 14 January; accepted 12 March 1999.

- Laughlin, R. B. Anomalous quantum Hall effect: an incompressible quantum fluid with fractional charge excitations. *Phys. Rev. Lett.* **50**, 1395–1398 (1983).
- Laughlin, R. B. Primitive and composite ground states in the fractional quantum Hall effect. *Surf. Sci.* **142**, 163–172 (1984).
- Prange, R. E. & Girvin, S. M. (eds) *The Quantum Hall Effect* (Springer, New York, 1987).
- de Picciotto, R. *et al.* Direct observation of a fractional charge. *Nature* **389**, 162–164 (1997).
- Saminadayer, L., Glatli, D. C., Jin, Y. & Etienne, B. Observation of the $e/3$ fractionally charged Laughlin quasiparticles. *Phys. Rev. Lett.* **79**, 2526–2529 (1997).
- Kane, C. L. & Fisher, M. P. A. Nonequilibrium noise and fractional charge in the quantum Hall effect. *Phys. Rev. Lett.* **72**, 724–727 (1994).
- Fendley, A., Ludwig, W. W. & Saleur, H. Exact nonequilibrium dc shot noise in Luttinger liquids and fractional quantum Hall devices. *Phys. Rev. Lett.* **75**, 2196–2199 (1995).
- de Chamon, C., Freed, D. E. & Wen, X. G. Tunneling and quantum shot noise in Luttinger liquids. *Phys. Rev. B* **51**, 2363–2378 (1995).
- Jain, J. K. Composite-fermion approach to the fractional quantum Hall effect. *Phys. Rev. Lett.* **63**, 199–202 (1989).
- Halperin, B. In *New Perspectives in Quantum Hall Effect* (eds Das Sarma, S. & Pinzuk, A.) 225–263 (Wiley, New York, 1997).
- Lesovik, G. B. Excess quantum shot noise in 2D ballistic point contacts. *JETP Lett.* **49**, 592–594 (1989).
- Reznikov, M., Heiblum, M., Shtrikman, H. & Mahalu, D. Temporal correlations of electrons: suppression of shot noise in a ballistic point contact. *Phys. Rev. Lett.* **75**, 3340–3334 (1995).
- Kumar, A., Saminadayer, L., Glatli, D. C., Jin, Y. & Etienne, B. Experimental test of the quantum shot noise reduction theory. *Phys. Rev. Lett.* **76**, 2778–2781 (1996).
- Reznikov, M. *et al.* Quantum shot noise. *Superlattices Microstruct.* **23**, 901–915 (1998).
- de Picciotto, R. Shot noise of non-interacting composite fermions. Preprint cond-mat/982221 at (xxx.lanl.gov) (1998).
- Sandler, N. P., de Chamon, C. & Fradkin, E. Noise measurements and fractional charge in fractional quantum Hall liquids. *Phys. Rev. B* (in the press); preprint cond-mat/9806335 at (xxx.lanl.gov) (1998).
- de Picciotto, R. *et al.* Direct observation of a fractional charge. *Phys. B* **249–251**, 395–400 (1998).
- Isakov, S., Martin, T. & Ouvry, S. Conductance and shot noise for particles with exclusion statistics. Preprint cond-mat/9811391 at (xxx.lanl.gov) (1998).

19. Fendley, P. & Saleur, H. Nonequilibrium dc noise in a Luttinger liquid with an impurity. *Phys. Rev. B* **54**, 10845–10854 (1996).
20. Martin, Th & Landauer, R. Wave packet approach to noise in multichannel mesoscopic systems. *Phys. Rev. B* **45**, 1742–1755 (1992).

Acknowledgements. We thank G. Bunin for manufacturing the transistors for the cooled preamplifier, and D. Mahalu for performing the e-beam lithography. This work was partly supported by the Israeli Science Foundation, the German-Israeli Foundation and the Israeli Ministry of Science.

Correspondence and requests for materials should be addressed to M.H. (e-mail: heiblum@wis. weizmann.ac.il)

Two types of avalanche behaviour in granular media

Adrian Daerr & Stéphane Douady

Laboratoire de Physique Statistique de l'ENS, 24 rue Lhomond, 75231 Paris CEDEX 05, France

The nature of the transition between static and flowing regimes in granular media^{1,2} provides a key to understanding their dynamics. When a pile of sand starts flowing, avalanches occur on its inclined free surface. Previously, studies³ of avalanches in granular media have considered the time series of avalanches in rotating drums⁴, or in piles continuously fed with material. Here we investigate single avalanches created by perturbing a static layer of glass beads on a rough inclined plane. We observe two distinct types of avalanche, with evidence for different underlying physical mechanisms. Perturbing a thin layer results in an avalanche propagating downhill and also laterally owing to collisions between neighbouring grains, causing triangular tracks; perturbing a thick layer results in an avalanche front that also propagates upwards, grains located uphill progressively tumbling down because of loss of support. The perturbation threshold for triggering an avalanche is found to decrease to zero at a critical slope. Our results may improve understanding of naturally occurring avalanches on snow slopes⁵ where triangular tracks are also observed.

The experiments are done on an inclined plane covered with velvet cloth. This surface is chosen so that the glass beads (180–300 μm in diameter), our granular material, have a larger friction with it than between themselves. A thin layer of grains can thus remain static on the plane up to a larger angle than if it were on a grain pile. We set the plane to an angle φ (larger than the pile angle φ_0) and pour glass beads abundantly at the top. The moving beads leave behind a static layer of uniform thickness $h(\varphi)$ (arrow leading to point a in Fig. 1; the geometry of the system is shown in Fig. 1 inset). This effect is explained by the variation of the friction of the successive grain layers with their distance to the surface of the inclined slope⁶: it is maximum for the bottom layer, and decreases continuously to the value for a thick pile. The top static layer is that which has a large enough friction coefficient on the underlying layers $\mu(h)$ to come to a stop. At inclination angle φ , the friction coefficient of this top static layer is then $\mu(h) = \tan\varphi$. The measurements $h(\varphi)$ of Fig. 1 thus give the variation of the coefficient of friction with depth⁷. We obtain a simple exponential decay as in ref. 7:

$$\mu(h) = \mu(\infty) + \Delta\mu[\exp(-h/h_0)] \quad (1)$$

with h_0 of the order of $2d$, where $d = 240 \mu\text{m}$ is taken as the mean diameter of the grains, and $\Delta\mu = \mu(0) - \mu(\infty)$ is the difference between the friction coefficient at the plane and that in the pile.

This remaining static layer is dynamically stable. If we add some beads onto this layer, they move downwards like a drop of liquid, leaving the layer unaffected. We can even jolt the experiment sharply: the whole layer starts to move but 'freezes' again very quickly.

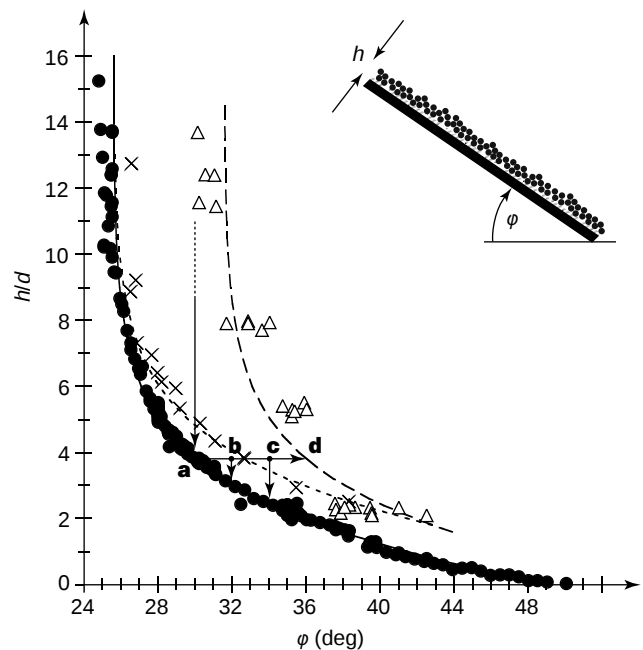


Figure 1 Stability diagram. Inset, the geometry of the experiment. Main figure, plot of h/d versus φ . The arrows sketch the course of an experiment. The filled circles show the measured thickness h of the layer left after a massive avalanche over the plane set at an angle φ (point a). The solid line is a fit according to equation (1). When these static layers are tilted they start flowing spontaneously at angles given by the open triangles (point d). The long-dashed line indicates the stability limit of a static layer. The hysteric region between the solid and the long-dashed line is separated by the crosses (and the short-dashed line). Below them (for example, point b) a finite disturbance will generate a triangular avalanche. Above them (for example, point c) it will result in an avalanche front which also propagates uphill. For preparation angles $\varphi > 34^\circ$, we observed only triangular avalanches (even the spontaneous ones). The precision of the measurements is smaller than the point size, and negligible compared to the physical fluctuations.

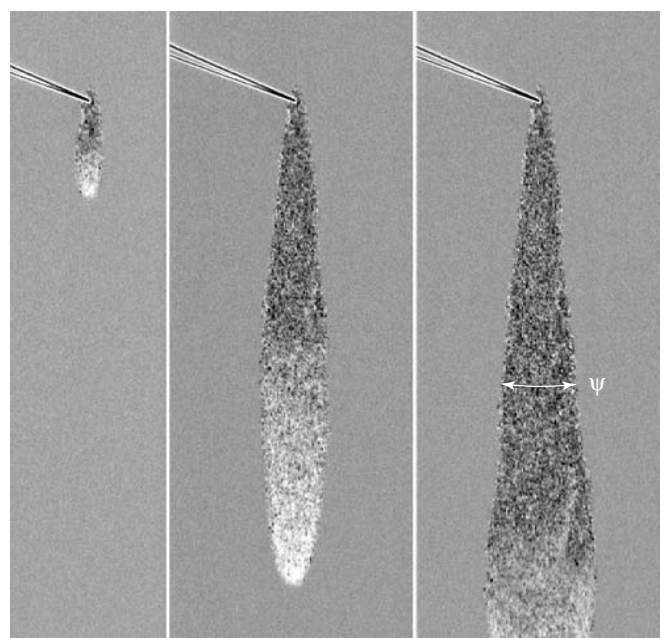


Figure 2 Evolution of a triangular avalanche ($\varphi = 30^\circ$, $\delta\varphi = 1.5^\circ$), showing the opening angle ψ . The time lapse between two images is 3.04 s. The pointed object is a pin used to trigger the avalanche, and indicates the origin of the avalanche.

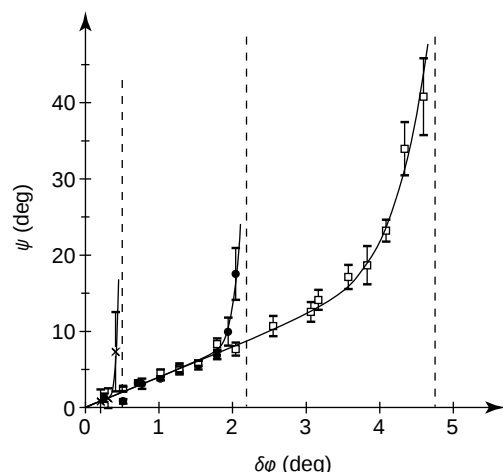


Figure 3 Opening angle ψ of triangular avalanches, as a function of the additional tilt $\delta\varphi$. The three sets of points correspond to $\varphi = 25^\circ$ (crosses), 28° (circles) and 32.5° (squares). For $\delta\varphi = 0$, we observe a drop, or a non-growing avalanche ($\psi = 0$). As $\delta\varphi$ increases, the opening angle first grows linearly, with a slope independent of φ . Just before the transition to the uphill propagation regime, (dashed lines, $\varphi = 0.5^\circ$, 2.25° and 4.75° , respectively), there is a sharp increase of ψ .

We then increase the tilting angle of the plane to $\varphi + \delta\varphi$ (from point a to point b in Fig. 1). For a small value of $\delta\varphi$, the layer remains perfectly static. But if we create a finite-sized disturbance (by setting some beads into motion or by pouring additional ones), an avalanche occurs. The height of the layer which is left behind an avalanche, $h(\varphi + \delta\varphi)$, is smaller than the initial height (Fig. 1). The avalanche, while running down, thus collects new material and becomes larger. Depending on both φ and $\delta\varphi$, we observe two types of propagation for the avalanches.

In the first type, for small $\delta\varphi$ or for thin layers (below the short-dashed line of Fig. 1, for example, point b), the avalanches grow laterally on their way down, leaving tracks of triangular shape. Figure 2 shows three successive pictures of such a 'triangular avalanche'. Although the total moving mass is increasing rapidly with time, the moving front reaches a constant velocity and thickness, the added rolling grains being spread behind the front (see Fig. 2). The lateral propagation is due only to a small number of grains which drive into motion some of their lateral neighbours. This process repeats, producing straight edges. The opening angle ψ of the resulting triangles (Fig. 2) increases at first proportionally to $\delta\psi$ (Fig. 3).

The second type of avalanche is obtained for large $\delta\varphi$ or for thick layers (above the short-dashed line of Fig. 1, for example, point c). In this case, the avalanche front propagates upwards; that is, the grains located uphill start to tumble down spontaneously due to the loss of their support. Contrary to the previous case, the whole layer will eventually flow. Figure 4 shows three successive pictures of such an uphill-propagating perturbation.

If we keep tilting the plane without doing any perturbation, the grain layer remains static up to a limiting value $\delta\varphi_s$ (point d in Fig. 1), where avalanches occur spontaneously. The angle hysteresis does not seem to depend much on the depth of the layer⁶, so the long-dashed line drawn in Fig. 1 has been chosen as the translation of $h(\varphi)$ by 6° as in ref. 6. The larger fluctuation observed here

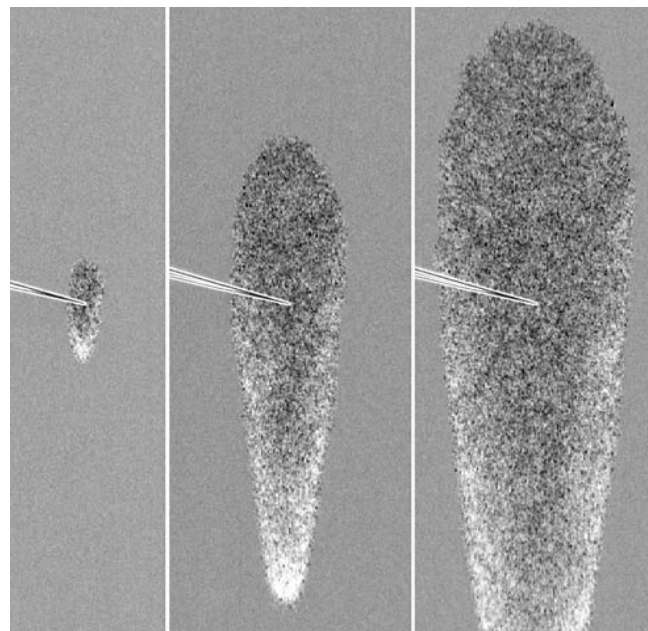


Figure 4 Three photographs of an uphill-propagating avalanche ($\varphi = 30^\circ$, $\delta\varphi = 2.5^\circ$). The point of the pin indicates where the avalanche was triggered: the time lapse between two images is 1.52 s. The upper front of the avalanche is now rounded, like a front propagating with different upward and sideward velocities. The uphill velocity jumps from zero to $\sim 5 \text{ cm s}^{-1}$ at the threshold, and then increases roughly linearly to 20 cm s^{-1} with increasing $\delta\varphi$. In the same interval, the velocity of the lower front increases linearly from 10 to 30 cm s^{-1} .

compared to ref. 6 could be due to the elasticity and roughness being larger for the material used here for the surface of the inclined plane (velvet) than that used in ref. 6 (glued beads).

To investigate this hysteresis between the static and dynamical angles, we did the following experiment: we pour a few grains very gently onto the surface of the layer, and record the minimum quantity of grains that is required, at a given $\delta\varphi$, to trigger an avalanche. The result (Fig. 5) shows a clear signature of a subcritical bifurcation: close to the initial slope φ , a considerable added height (of more than twice the layer thickness) is needed to get an avalanche (or a drop for $\delta\varphi = 0$) started. Less and less perturbation is needed as $\delta\varphi$ increases, until finally, it can become as small as one grain.

We note that the tilt angle for which only one grain starts an avalanche is smaller than the one for which a spontaneous avalanche occurs. This comes from the fact that the falling grain is already in motion, while observations show that spontaneous avalanches occur when a previously motionless grain spontaneously rolls down. This sensitivity to the stability of one surface grain is the source of the large dispersion measured for the static limit (Fig. 1) when the experiment is repeated: the surface has a fluctuating roughness, with some grains more unstable than others.

The transition from lateral and downhill propagation to uphill propagation could at first be interpreted as the point where the opening angle of triangular avalanches exceeds 180° (similar to the subsonic/supersonic transition or convective/absolute instability). But measuring $\psi(\delta\varphi)$ for several φ (Fig. 3) showed that triangular avalanches close to the threshold for upward motion still have small opening angles. The transition to upward propagation is not linked either to any particular value of $\delta\varphi$: the threshold on $\delta\varphi$ for upward propagation increases with φ . But inside the triangular avalanche domain, in contrast, the $\psi(\delta\varphi)$ curves for different φ can be superposed, which means that there is no dependence on φ . These observations indicate that the physical mechanisms underlying the lateral propagation of the triangular avalanches are

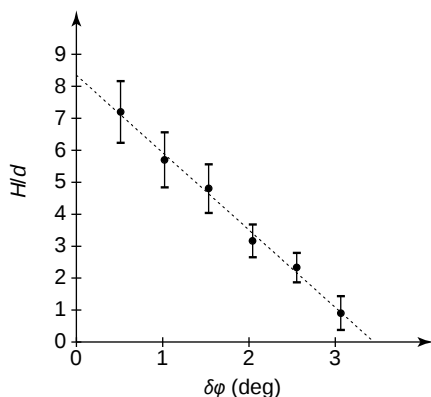


Figure 5 Avalanche threshold. The initial angle is $\varphi = 32^\circ$; the height of the minimum perturbation triggering an avalanche is plotted as a function of the additional tilt $\delta\varphi$. We use an inclined tube to release from a very small height additional glass beads practically one by one at one spot on the surface⁸. As soon as an avalanche has occurred, the delivery is stopped and the added mass m is measured. We calculate an equivalent added height $H = [(3/\pi)[\tan^2(\beta)]m/\rho]^{1/3}$ as the beads form a cone of slope $\beta \approx 30^\circ$ and mean density $\rho = 1.48 \text{ g cm}^{-3}$ on top of the existing layer. This height decreases roughly linearly with $\delta\varphi$, as indicated by the short-dashed line. The error bars are the statistical error over ten measurements.

different in nature from those for upward propagation.

Careful observation of the motions of grains shows that these reach their limiting velocity very rapidly and go straight down the slope. The lateral spread of the triangular avalanches can thus be interpreted as due to the friction exerted by the rolling grains on their lateral neighbours being sufficient to set some of them into motion. On the other hand, in upward propagation, the grains start to roll down spontaneously because the grains located below them are already gone. The threshold of this uphill motion is compatible with the condition that the thickness difference between the initial layer and the new static layer is larger than a typical height, taken as one particle diameter for the short-dashed line in Fig. 1.

We observed the same two types of behaviour with other materials (crushed walnut shells and Sinai sand). We can thus think of comparing these avalanches with other types of avalanche (of snow or rocks) occurring in a geological context. In these cases, the triggering, amplification and velocity of the avalanches are of practical importance. Comparisons can already be made with some snow avalanches, where triangular tracks are observed at their origin⁵. In our experiment we have observed triangular avalanches exclusively in the case where a thin layer is located over a surface creating a strong friction. In all the other cases, the avalanche front propagates uphill. This suggests that the variation of friction coefficient across the layer also plays an important role in snow avalanches. □

Received 30 November 1998; accepted 11 March 1999.

1. Reynolds, O. On the dilatancy of media composed of rigid particles in contact. *Phil. Mag.* **20**, 469–481 (1885).
2. Bagnolds, R. A. The shearing and dilatation of dry sand and the 'singing' mechanism. *Proc. R. Soc. Lond. A* **295**, 219–232 (1966).
3. Rajchenbach, J. in *Physics of Dry Granular Media* (eds Hermann H., Hovi, J.-P. & Luding, S.) 421–440 (Kluwer, Dordrecht, 1998).
4. Caponieri, M., Douady, S., Fauve, S. & Laroche, C. in *Mobile Particulate Systems* (eds Guazzelli, E. & Oger, L.) 331–366 (Kluwer, Dordrecht, 1995).
5. McClung, D. *Avalanche Handbook* (Mountaineers, Seattle, 1993).
6. Pouliquen, O. & Renault, N. Onset of granular flows on an inclined rough surface: dilatancy effects. *J. Phys. II France* **6**, 923–935 (1996).
7. Pouliquen, O. Scaling laws in granular flows down rough inclined planes. *Phys. Fluids* (in the press).
8. Held, G. A. *et al.* Experimental study of critical-mass fluctuations in an evolving sandpile. *Phys. Rev. Lett.* **65**, 1120–1123 (1990).

Acknowledgements. This work was supported by the University Paris VII through BQR97.

Correspondence and requests for materials should be addressed to A.D. (e-mail: daerr@physique.ens.fr).

Desorption–ionization mass spectrometry on porous silicon

Jing Wei*, Jillian M. Buriak† & Gary Siuzdak*

* Departments of Molecular Biology and Chemistry, The Scripps Research Institute, BCC157, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

† Department of Chemistry, Purdue University, West Lafayette, Indiana 47907-1393, USA

Desorption mass spectrometry has undergone significant improvements since the original experiments were performed more than 90 years ago¹. The most dramatic change occurred in the early 1980s with the introduction of an organic matrix^{2–4} to transfer energy to the analyte. This reduces ion fragmentation but also introduces background ions from the matrix. Here we describe a matrix-free strategy for biomolecular mass spectrometry based on pulsed-laser desorption–ionization from a porous silicon⁵ surface. Our method uses porous silicon to trap analytes deposited on the surface, and laser irradiation to vaporize and ionize them. We show that the method works at femtomole and attomole levels of analyte, and induces little or no fragmentation, in contrast to what is typically observed with other such approaches^{6–11}. The ability to perform these measurements without a matrix^{3,4,12,13} also makes it more amenable to small-molecule analysis. Chemical¹⁴ and structural¹⁵ modification of the porous silicon has enabled optimization of the ionization characteristics of the surface. Our technique offers good sensitivity as well as compatibility with silicon-based microfluidics and microchip technologies.

The broad success of matrix-assisted desorption/ionization is related to the ability of the matrix to incorporate and transfer energy to the analyte^{3,4,12,16}. For instance, in matrix-assisted laser desorption/ionization (MALDI)⁴, the analyte is typically dissolved into a solid ultraviolet-absorbing organic acid matrix which vaporizes upon pulsed-laser radiation, carrying with it the analyte¹⁶. Direct desorption/ionization without matrix has been extensively studied on a variety of surfaces^{6–11} but has not been widely used because of the rapid molecular degradation that is usually observed upon direct exposure to laser radiation. However, just as secondary ion mass spectrometry (SIMS) has had a profound effect on surface science¹⁷, the utility of direct laser desorption/ionization for biomolecular analysis could be highly beneficial owing to dramatically simplified sample preparation, elimination of matrix background ions, and other advantages we describe here.

Our matrix-free desorption approach uses the porous silicon surface to trap the analyte molecules and, because of its high absorptivity in the ultraviolet¹⁸, act as an energy receptacle for the laser radiation. The experimental protocol for desorption/ionization on silicon (DIOS) (see Methods and Fig. 1) involves the generation of porous silicon from flat crystalline silicon by using a simple galvanostatic etching procedure¹⁵. A micrometres-thick porous layer with a nanocrystalline architecture is produced that exhibits bright photoluminescence upon exposure to ultraviolet light^{5,19}. The thickness, morphology, porosity, resistivity and other characteristics of the material are readily modulated through the choice of silicon-wafer precursor and etching conditions (see Methods). Freshly etched porous silicon surfaces are hydrophobic owing to the metastable, silicon-hydride termination, but through Lewis-acid-mediated¹⁴ or light-promoted hydrosilylation reactions²⁰, the surface can be easily stabilized and functionalized as required. Because of the high stability of the hydrophobic, hydrosilylated surfaces to aqueous media, these surfaces can be reused repeatedly (see Methods) with little degradation^{14,20}. Photopatterning of the

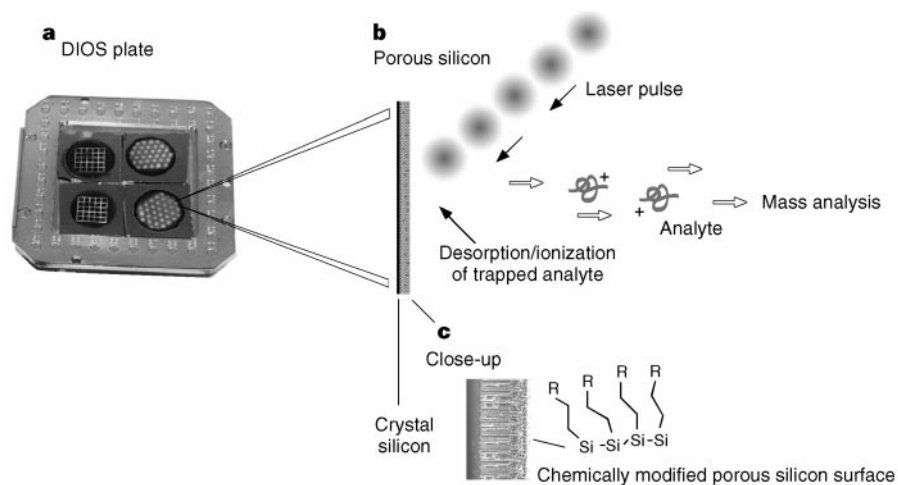


Figure 1 Experimental configuration for the DIOS-MS experiments. **a**, Four porous silicon plates are placed on a MALDI plate. Each of the porous silicon plates contains photopatterned spots or grids prepared through illumination of n-type silicon with a 300-W tungsten filament through a mask and an $f/50$ reducing lens. **b**, The silicon-based laser desorption/ionization process, in which

the sample is placed on the porous silicon plate and allowed to dry, followed by laser-induced desorption/ionization mass spectrometry. **c**, Cross-section of porous silicon, and the surface functionalities after hydrosilylation; R represents phenyl or alkyl chains.

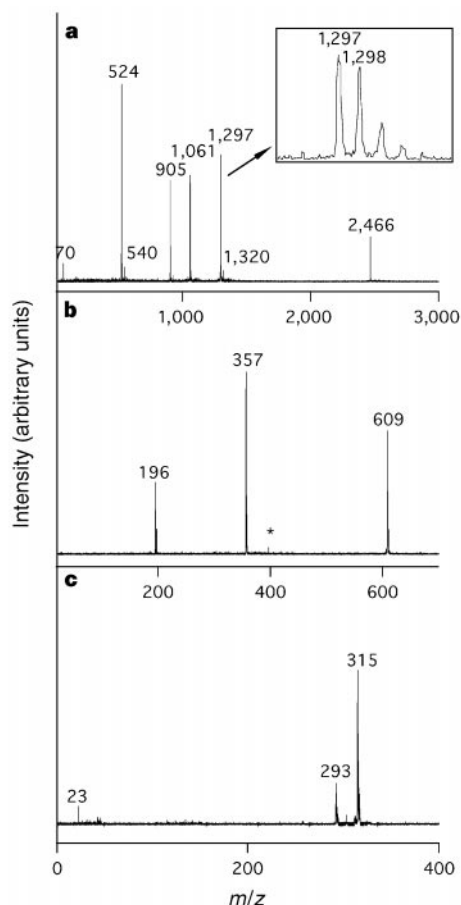


Figure 2 Examples of mass spectral data obtained with DIOS. **a**, The DIOS mass spectrum of a mixture of five peptides (2 pmol each), including a four-residue peptide (MRFA in single-letter amino-acid code) at m/z 524, des-arg-bradykinin (m/z 905), bradykinin (m/z 1,061), angiotensin (m/z 1,297), and adrenocorticotrophic hormone (m/z 2,466). The small peaks at m/z 540 and m/z 1,320 are oxidized MRFA and a sodium adduct of angiotensin, respectively. The signal of m/z 70 corresponds to a surface background ion (possibility $C_5H_{10}^+$). Insert, spectrum

showing the isotopes of angiotensin and that the resolution is not affected by the porous silicon surface. **b**, The DIOS mass spectrum of a mixture of three small molecules, including caffeine (m/z 196), an antiviral drug WIN (m/z 357) and reserpine (m/z 609) (1 pmol each). A small signal (asterisk) was an impurity from caffeine. **c**, The DIOS mass spectrum of 10 pmol N -octyl β -D-glucopyranoside (m/z 293) and its sodium adduct ion (m/z 315). The sodium ion (m/z 23) itself was also detected.

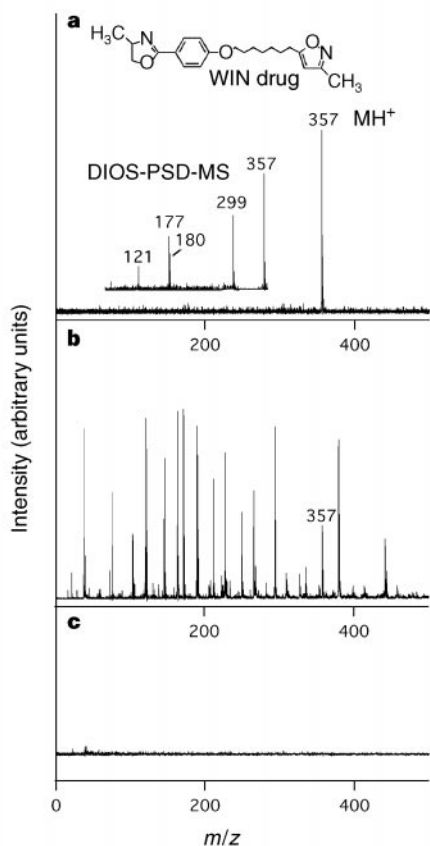


Figure 3 Analysis of WIN antiviral drug (500 fmol) using different desorption/ionization techniques. **a**, DIOS-MS spectrum of WIN. Accurate mass measurements were obtained on WIN with a time-of-flight reflectron instrument, typically to within 10 p.p.m. (the limit of accuracy of this instrument in this mass range). The inset spectrum represents post-source decay fragmentation measurements on WIN. These results are consistent with results from electrospray ionization tandem mass spectrometry experiments. **b**, MALDI-MS of same amount of antiviral drug using α -cyano-4-hydroxycinnamic matrix. **c**, No signal was obtained from any of the compounds tested using laser desorption mass spectrometry off the gold MALDI plate.

surfaces is possible and has been used galvanostatically to etch 5×5 -well plates in a 1.1-cm^2 area on n-type silicon, allowing for the analysis of 25 samples in series²¹ (Fig. 1). We used surface photopatterning to identify where on the uniform porous silicon surface the sample had been placed.

The porous silicon surfaces were tested with a broad range of compounds. Examples of DIOS mass spectrometry (DIOS-MS) on peptides, caffeine, an antiviral drug molecule (WIN), reserpine and *N*-octyl β -D-glucopyranoside are shown in Fig. 2. Over thirty other compounds ranging in size from 150 to 12,000 daltons, including carbohydrates, peptides, glycolipids, natural products and small drug molecules, were tested and their molecular ion was observed with little or no fragmentation. Although MALDI-MS analysis is also possible for small molecules²² and matrix suppression can be achieved under certain circumstances²³, matrix interference presents a real limitation (Fig. 3) that a matrix-less technique such as DIOS does not encounter.

DIOS has some unique features that make it useful for a variety of biomolecules. Peptides generate a good signal from the deposition of 700 attomol of material (also the limit of detection in our MALDI instrument on this compound) and allow the analyte to be analysed even in a saturated salt solution. For example, spectra from the peptide des-arg-bradykinin (Fig. 4) were easily obtained from

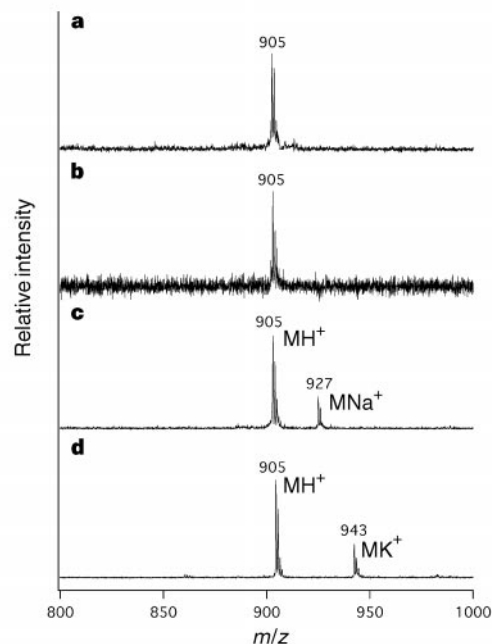


Figure 4 DIOS mass spectra of des-arg-bradykinin using small quantities of sample and in the presence of salt. **a–d**, DIOS mass spectra with: **a**, 7 fmol; **b**, 700 attmol; **c**, 2 pmol in the presence of a 2 M NaCl; and **d**, 2 pmol in a saturated solution of K_3PO_4 buffer solution. All spectra shown were generated from an n-type mesoporous silicon surface modified with an ethyl phenyl termination; they are the average of 128 shots with a nitrogen laser (337 nm) at an intensity, typical of MALDI experiments, of 2 to 50 μJ per pulse.

saturated K_3PO_4 , 2.0 M NaCl, and 2.0 M Tris solutions, although higher laser intensities were required for these analyses. No silicon-containing adducts were ever detected in DIOS-MS spectra which might interfere with these analyses, indicating the inert nature of the porous silicon material. In addition, the resolution obtained in the analysis of compounds (Fig. 2) from DIOS was identical to that in MALDI analyses, as were post-source decay fragmentation measurements made on peptides and small molecules (Fig. 3). However, the DIOS post-source decay small-molecule measurements would ordinarily be impossible to perform with a MALDI reflectron instrument because of matrix interference.

The observation of a variety of intact ions directly emanating from these surfaces suggests that porous silicon has properties such as high surface area²⁴ and strong ultraviolet absorption¹⁸ that may enhance pulsed laser desorption and ionization. The porous structure on the surface is essential for desorption and ionization, because control experiments done with glass, with air-oxidized single-crystal silicon of (100) orientation, with 0.25-mm-thick, 60 Å-pore silica-gel thin-layer chromatography plates, or with gold MALDI plates (Fig. 3) gave no significant ion signal. A variety of etching conditions²⁵ were used to produce microporous (≤ 2 nm pore sizes) and mesoporous (2–50 nm pore sizes) silicon²⁴. Both n-type mesoporous samples and p-type micro- or mesoporous samples were effective in generating signals, but the surfaces with smaller pore sizes typically gave a more intense ion signal. To understand the ionization mechanism better, we used the DIOS approach on four porous silicon surfaces, each containing different surface modifications, including hydrogen (native), dodecyl, ethyl phenyl ($-\text{CH}_2\text{CH}_2\text{C}_6\text{H}_5$), and oxide. Of the porous silicon surfaces tested, most could effectively generate ions. However, the more hydrophobic surfaces, and in particular an ethylphenyl-terminated surface, typically gave better signals for the same quantity of analyte from an aqueous medium¹⁴. Dissolving the sample in methanol/ H_2O at a 1:1 v/v ratio also provided a stronger signal, indicating that

analyte penetration into the porous silicon is important. The energy for analyte release from the surface may be transferred from silicon to the trapped analyte through vibrational pathways or as a result of the rapid heating of porous silicon producing H_2 (ref. 26) that releases the analyte. Alternatively, as porous silicon absorbs hydrocarbons from air or while under vacuum^{27,28}, rapid heating/vaporization of trapped hydrocarbon contaminants or solvent molecules may augment the vaporization and ionization of the analyte embedded in the porous silicon. The observation of DIOS-MS background ions ($m/z < 100$) at high laser intensities, consistent with aliphatic hydrocarbons, indicates that they may play a role in desorption/ionization.

The potential application of DIOS-MS is broad. The surface activity could be changed through derivatization with receptors²⁹, aiding the identification of ligands. In addition, the enhanced signal obtained from the ethyl phenyl surface suggests that significant improvements can be made to DIOS-MS through further investigation of surface modifications. Most existing MALDI mass spectrometers could be used to perform DIOS simply by making modifications to the MALDI sample plate (Fig. 1). Moreover, porous silicon material is inexpensive and its production simple. Porous silicon can be readily integrated with existing silicon-based technology¹⁵, allowing, for example, its application into miniaturized chip, microfluidic chemical reactors that are lithographically etched into crystalline silicon wafers. For instance, porous silicon features as small as 20 μm and 100 nm can be produced through standard optical techniques²¹ and ion implantation³⁰, respectively. In short, as a new desorption/ionization approach, DIOS-MS offers sensitivity, high tolerance to contaminants, no matrix interference and, because the surface properties of porous silicon can easily be tailored, more control over biomolecular mass spectrometry. $\square$

Methods

Effective porous silicon samples for DIOS could be prepared from either n- or p-type silicon according to these general procedures. For n-type: P-doped, (100) orientation, 0.65 Ωcm resistivity Si wafers were etched for 1–3 min using +71 mA cm^{-2} current density with illumination by a 300-W tungsten filament bulb in a 1:1 solution of ethanol/49% HF (aq). For p-type: B-doped, (100) orientation, 0.01 Ωcm resistivity Si wafers were etched at 37 mA cm^{-2} current density in the dark for 3 h in a 1:1 solution of ethanol/49% HF (aq).

DIOS, MALDI and laser desorption (off a gold surface) experiments were performed on a Voyager DE-STR, time-of-flight mass spectrometer (PerSeptive Biosystems) using a pulsed nitrogen laser (Laser Science) operated at 337 nm. Once formed, ions were accelerated into the time-of-flight mass analyser with a voltage of 20 kV. Porous silicon surfaces for DIOS analysis were mounted onto the MALDI probe by adhesive tape. Analytes were dissolved in a deionized H_2O or H_2O /methanol mixture at concentrations typically ranging from 0.001 to 10.0 μM . Aliquots (0.5–1.0 μl ; corresponding to 0.5 femtomol to 100 picomol analyte) of solution were deposited onto the porous surfaces and allowed to dry before DIOS mass-spectrometry analysis. After multiple usage, the porous silicon surface-regeneration procedure necessitated rinsing surfaces with deionized H_2O and methanol sequentially, and finally immersing the surface overnight in a 1:2 v/v methanol/ H_2O mixture. Surfaces were rinsed with deionized H_2O then with methanol and allowed to dry before applying the analyte.

Received 17 November 1998; accepted 1 March 1999.

- Thomson, J. J. Rays of positive electricity. *Phil. Mag.* **20**, 752–767 (1910).
- Liu, L. K., Busch, K. L. & Cooks, R. G. Matrix-assisted secondary ion mass spectra of biological compounds. *Anal. Chem.* **53**, 109–113 (1981).
- Barber, M., Bordoli, R. S., Sedgwick, R. D. & Tyler, A. N. Fast-atom bombardment of solids as an ion source in mass spectrometry. *Nature* **293**, 270–275 (1981).
- Karas, M. & Hillenkamp, F. Laser desorption/ionization of proteins with molecular mass exceeding 10,000 daltons. *Anal. Chem.* **60**, 2299–2301 (1988).
- Canham, L. T. Silicon quantum wire array fabrication by electrochemical and chemical dissolution of wafers. *Appl. Phys. Lett.* **57**, 1046–1048 (1990).
- Zenobi, R. Laser-assisted mass spectrometry. *Chimia* **51**, 801–803 (1997).
- Zhan, Q., Wright, S. J. & Zenobi, R. Laser desorption substrate effects. *J. Am. Soc. Mass Spectr.* **8**, 525–531 (1997).
- Hrubowchak, D. M., Ervin, M. H., Wood, M. C. & Winograd, N. Detection of biomolecules on surfaces using ion-beam-induced desorption and multiphoton resonance ionization. *Anal. Chem.* **63**, 1947–1953 (1991).

- Varakin, V. N., Lunchev, V. A. & Simonov, A. P. Ultraviolet-laser chemistry of adsorbed dimethylcadmium molecules. *High En. Chem.* **28**, 406–411 (1994).
- Wang, S. L., Ledingham, K. W. D., Jia, W. J. & Singhal, R. P. Studies of silicon nitride (Si_3N_4) using laser ablation mass spectrometry. *Appl. Surf. Sci.* **93**, 205–210 (1996).
- Posthumus, M. A., Kistemaker, P. G., Meuzelaar, H. L. C. & Ten Noever de Brauw, M. C. Laser desorption–mass spectrometry of polar nonvolatile bio-organic molecules. *Anal. Chem.* **50**, 985–991 (1978).
- Macfarlane, R. D. & Torgerson, D. F. Californium-252 plasma desorption mass spectroscopy. *Science* **191**, 920–925 (1976).
- Siuzdak, G. in *Mass Spectrometry for Biotechnology* 162 (Academic, San Diego, 1996).
- Buriak, J. M. & Allen, M. J. Lewis acid mediated functionalization of porous silicon with substituted alkenes and alkynes. *J. Am. Chem. Soc.* **120**, 1339–1340 (1998).
- Cullis, A. G., Canham, L. T. & Calcott, P. D. J. The structural and luminescence properties of porous silicon. *J. Appl. Phys.* **82**, 909–965 (1997).
- Hillenkamp, F., Karas, M., Beavis, R. C. & Chait, B. T. Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. *Anal. Chem.* **63**, A1193–A1202 (1991).
- Benninghoven, A., Rüdenauer, F. G. & Werner, H. W. (eds) *Secondary Ion Mass Spectrometry* 1227 (Wiley, New York, 1987).
- Amato, G. & Rosenbauer, M. in *Optoelectronic Properties of Semiconductors and Superlattices* (eds Amato, G., Delerue, C. & Bardeleben, H.-J.) 3–52 (Gordon and Breach, Amsterdam, 1997).
- Sailor, M. J. & Lee, E. J. Surface chemistry of luminescent silicon nanocrystallites. *Adv. Mater.* **9**, 783–793 (1997).
- Stewart, M. P. & Buriak, J. M. Photopatterned hydrosilylation on porous silicon. *Angew. Chem. Int. Edn* **37**, 3257–3261 (1998).
- Doan, V. V. & Sailor, M. J. Photolithographic fabrication of micron-dimension porous Si structures exhibiting visible luminescence. *Appl. Phys. Lett.* **60**, 619–620 (1992).
- Lidgard, R. & Duncan, M. W. Utility of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry for the analysis of low molecular weight compounds. *Rapid Commun. Mass Spectr.* **9**, 128–132 (1995).
- Knochenmuss, R., Dubois, F., Dale, M. J. & Zenobi, R. The matrix suppression effect and ionization mechanisms in matrix-assisted laser desorption/ionization. *Rapid Commun. Mass Spectr.* **10**, 871–877 (1996).
- Canham, L. T. in *Properties of Porous Silicon* (ed. Canham, L. T.) 83–88 (Institution of Electrical Engineers, London, 1997).
- Hérino, R. in *Properties of Porous Silicon* (ed. Canham, L. T.) 89–96 (Institution of Electrical Engineers, London, 1997).
- Okada, Y. H., Kato, F., Tsuboi, T. & Sakka, T. J. Changes in the environment of hydrogen in porous silicon with thermal annealing. *Electrochem. Soc.* **145**, 2439–2444 (1998).
- Canham, L. T. in *Properties of Porous Silicon* (ed. Canham, L. T.) 154–157 (Institution of Electrical Engineers, London, 1997).
- Canham, L. T. in *Properties of Porous Silicon* (ed. Canham, L. T.) 44–50 (Institution of Electrical Engineers, London, 1997).
- O'Donnell, M. J., Tang, K., Koster, H. & Smith, C. L. High density, covalent attachment of DNA to silicon wafers for analysis by MALDI-TOF mass spectrometry. *Anal. Chem.* **69**, 2438–2443 (1997).
- Schmuki, P., Erickson, L. E. & Lockwood, D. J. Light emitting micropatterns of porous Si created at surface defects. *Phys. Rev. Lett.* **80**, 4060–4063 (1998).

Acknowledgements. We thank J. Boydston, Z. Shen, R. G. Cooks and M. Duncan for their comments, T. Hollenbeck for suggesting the acronym DIOS, K. Harris for his initial analyses of porous silicon surfaces, and M. P. Stewart and T. Geders for preparing oxidized and p-type porous silicon samples. G.S. is grateful for partial funding by an NIH grant; J.M.B. thanks Purdue University for support and the Camille and Henry Dreyfus Foundation for a New Faculty Award.

Correspondence and requests for materials should be addressed to J.M.B. and G.S. (e-mail: buriak@purdue.edu and siuzdak@scripps.edu).

Growing range of correlated motion in a polymer melt on cooling towards the glass transition

Christoph Bennemann*†, Claudio Donati*†, Jörg Baschnagel‡ & Sharon C. Glotzer*

* Center for Theoretical and Computational Materials Science, and Polymers Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

† Dipartimento di Fisica e Istituto Nazionale per la Fisica della Materia, Università di Roma 'La Sapienza', Piazzale Aldo Moro 2, I-00185 Roma, Italy

‡ Institut für Physik, Johannes Gutenberg Universität, Staudinger Weg 7, 55099 Mainz, Germany

Many liquids cooled to low temperatures form glasses (amorphous solids) instead of crystals. As the glass transition is approached, molecules become localized and relaxation times increase by many orders of magnitude¹. Many features of this 'slowing down' are reasonably well described² by the mode-coupling theory of supercooled liquids³. The ideal form of this theory predicts a dynamical critical temperature T_c at which the

molecules become permanently trapped in the 'cage' formed by their neighbours, and vitrification occurs. Although there is no sharp transition, because molecules do eventually escape their cage, its signature can still be observed in real and simulated liquids. Unlike conventional critical phenomena (such as the behaviour at the liquid–gas critical point), the mode-coupling

transition is not accompanied by a diverging static correlation length. But simulation^{4–10} and experiment^{11,12} show that liquids are dynamically heterogeneous, suggesting the possibility of a relevant 'dynamical' length scale characterizing the glass transition. Here we use computer simulations to investigate a melt of short, unentangled polymer chains over a range of temperatures for which the mode-coupling theory remains valid. We find that although density fluctuations remain short-ranged, spatial correlations between monomer displacements become long-ranged as T_c is approached on cooling. In this way, we identify a growing dynamical correlation length, and a corresponding order parameter, associated with the glass transition. This finding suggests a possible connection between well established concepts in critical phenomena and the dynamics of glass-forming liquids.

We consider a polymer melt consisting of chains of M identical monomers. Let the position of each monomer i at a time t be denoted $\mathbf{r}_i(t)$. As a measure of the local motion, we introduce the scalar displacement $\mu_i(t, \Delta t) = |\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)|$ over some interval of time Δt , starting from a reference time t . We examine the spatial correlations of these displacements by modifying the definition of the usual static density–density correlation function $G(\mathbf{r})$ (ref. 13) so that the contribution of a monomer i to the correlation function is weighted by μ_i . That is, we define a 'displacement–displacement' correlation function^{14,15},

$$G_u(\mathbf{r}, \Delta t) = \int d\mathbf{r}' \langle [\mu(\mathbf{r}' + \mathbf{r}, t, \Delta t) - \langle \mu \rangle][\mu(\mathbf{r}', t, \Delta t) - \langle \mu \rangle] \rangle$$

where

$$u(\mathbf{r}, t, \Delta t) = \sum_{i=1}^N \mu_i(t, \Delta t) \delta(\mathbf{r} - \mathbf{r}_i(t))$$

Here $\langle \dots \rangle$ indicates an ensemble average, and N is the total number of monomers present in a particular configuration. $G_u(\mathbf{r}, \Delta t)$ measures the correlations of fluctuations of local displacements away from their average value $\langle \mu \rangle \equiv \langle \mu(\mathbf{r}, t, \Delta t) \rangle$. As the melt is isotropic, $G(\mathbf{r})$ and $G_u(\mathbf{r}, \Delta t)$ only depend on $r = |\mathbf{r}|$. We define

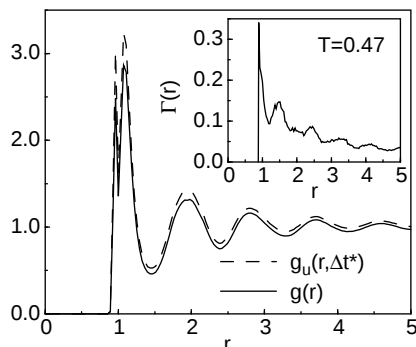


Figure 1 Displacement–displacement correlation function $g_u(r, \Delta t)$ and pair correlation function $g(r)$ plotted versus r at $T = 0.47$ and $\Delta t = 277.76$. Inset: $\Gamma(r) = [g_u(r, \Delta t)/g(r)] - 1$, plotted versus r . Note that if the displacement of monomers were spatially uncorrelated then $g_u(r, \Delta t)$ and $g(r)$ would be identical for all r . Hence, it is deviations of $g_u(r, \Delta t)$ from $g(r)$ that demonstrate spatial displacement correlations in excess of those expected from $g(r)$ alone. Results were obtained by a molecular dynamics simulation of a polymer melt consisting of about 120 chains of 10 monomers each. Monomers interact via a Lennard-Jones potential, and nearest-neighbour monomers along a chain are connected by a FENE anharmonic spring potential^{19–21}. Length and temperature scales are measured in Lennard-Jones units. Simulations are carried out at constant pressure ($P = 1$) and in a temperature range (T) from 0.70–0.46, corresponding to densities of $0.91 \leq n \leq 1.04$. Each state point is fully equilibrated before all measurements. For each state point, all quantities are averaged over 150–200 independent time origins. This melt is well described by ideal mode-coupling theory with $T_c = 0.450 \pm 0.005$ and $n_c = 1.042 \pm 0.010$ (refs 19–21).

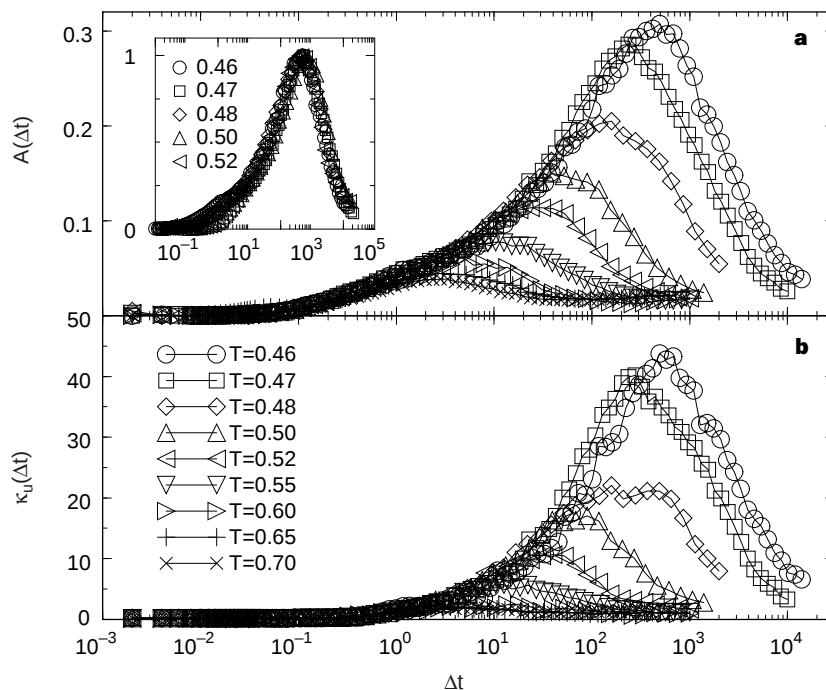


Figure 2 Time dependence of the 'excess' correlation $A(\Delta t)$ and the susceptibility $\kappa_u(\Delta t)$. **a**, $A(\Delta t)$ plotted versus Δt for different T . Inset: $A(\Delta t)/A(\Delta t^*)$ plotted versus $\Delta t/\Delta t^*$ for lowest five temperatures, showing that the data collapse onto a single

master curve over several decades. **b**, $\kappa_u(\Delta t)$ plotted versus Δt for the same T as in **a**. Although not as clean, $\kappa_u(\Delta t)$ appears to show the same data collapse as that for $A(\Delta t)$ (inset).

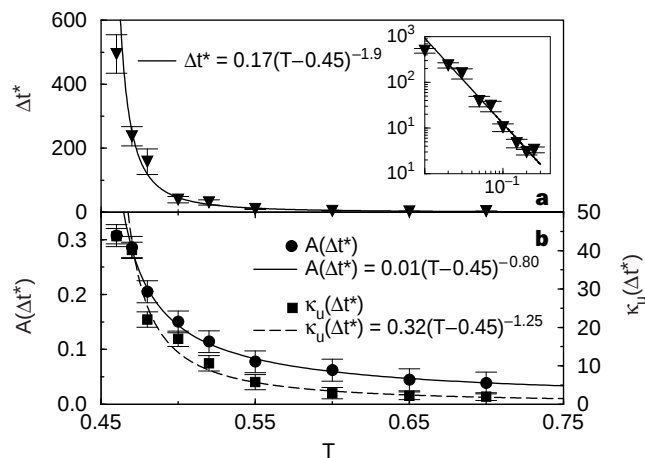


Figure 3 Temperature dependence of the 'excess' correlation $A(\Delta t)$ and the 'susceptibility' $\kappa_u(\Delta t)$. **a**, Time Δt^* when $A(\Delta t)$ is largest, plotted versus T . The solid curve is a power-law fit, $\Delta t^* \sim (T - T_c)^{-\gamma}$, with $T_c = 0.450 \pm 0.005$ and $\gamma = 1.87 \pm 0.15$. Inset: same data and fit, plotted log-log versus $T - 0.450$. **b**, κ_u and A evaluated at Δt^* , plotted versus T . The curves are power-law fits to the data obtained with $T_c = 0.450$, as indicated. Excellent fits are obtained for a small range of $T_c = 0.450 \pm 0.005$, and corresponding $\gamma = 0.80 \pm 0.07$ for A , and $\gamma = 1.25 \pm 0.11$ for κ . Note that the data point for $T = 0.46$ is excluded from all fits. All fits are made by subtracting T_c from T , taking the log of the ordinate and the abscissa, and performing a linear regression of the logged data to obtain the exponent and prefactor of the power law for that value of T_c . Error bars on the exponents are obtained by varying T_c between the lower and upper bounds quoted here.

the 'total displacement' $U(t, \Delta t) = \int d\mathbf{r} u(\mathbf{r}, t, \Delta t)$ and its ensemble average $\langle U \rangle = \langle U(t, \Delta t) \rangle$. In a constant- N ensemble, both $\langle u \rangle$ and $\langle U \rangle$ are readily evaluated from the mean displacement $\bar{\mu} \equiv \langle N^{-1} \sum_{i=1}^N \mu_i(t, \Delta t) \rangle$: $\langle u \rangle = \bar{\mu}n$ and $\langle U \rangle = \bar{\mu}N$, where $n (= N/V)$ is the density. In equilibrium, $\langle u \rangle$, $\langle U \rangle$, $\bar{\mu}$ and $G_u(\mathbf{r}, \Delta t)$ do not depend on t , but they retain a dependence on Δt which we address below.

$G_u(\mathbf{r}, \Delta t)$ can be written so as to identify a spatial correlation function $g_u(\mathbf{r}, \Delta t)$, analogous to the pair correlation $g(\mathbf{r})$ conventionally used to characterize the structure of a liquid

$$G_u(\mathbf{r}, \Delta t) = N\bar{\mu}^2\delta(\mathbf{r}) + \langle u \rangle \langle U \rangle [g_u(\mathbf{r}, \Delta t) - 1]$$

where

$$g_u(\mathbf{r}, \Delta t) = \frac{1}{\langle u \rangle \langle U \rangle} \left\langle \sum_{i=1}^N \sum_{j=1}^N \mu_i(t, \Delta t) \mu_j(t, \Delta t) \delta(\mathbf{r} + \mathbf{r}_j(t) - \mathbf{r}_i(t)) \right\rangle$$

The mean-square displacement $\bar{\mu}^2 \equiv \langle \frac{1}{N} \sum_{i=1}^N \mu_i^2(t, \Delta t) \rangle$ also depends on Δt . The fluctuations of U are related to the volume integral of $G_u(\mathbf{r}, \Delta t)$:

$$\langle [U - \langle U \rangle]^2 \rangle = \int d\mathbf{r} G_u(\mathbf{r}, \Delta t) \equiv \langle U \rangle \langle u \rangle kT \kappa_u \quad (1)$$

Here we have defined the quantity κ_u in analogy to the isothermal compressibility κ , which is proportional to the volume integral of $G(\mathbf{r})$ (ref. 16). Recall that near a conventional critical point where κ diverges, long-range behaviour is observed for $G(\mathbf{r})$ that is associated with the occurrence of macroscopic density fluctuations¹⁶. It is thus possible to determine the large- r behaviour of $G_u(\mathbf{r})$ from the fluctuations of U , just as the large- r behaviour of $G(\mathbf{r})$ can be determined from the fluctuations of N (ref. 16).

In the present polymer melt, as in most glass-forming liquids^{17,18}, we observe no growing static compressibility, and thus no long-range correlations in the monomer positions, on cooling, despite an increase of relaxation times of 3–4 orders of magnitude over the range of T investigated^{19–21}. However, a growing range of correlation

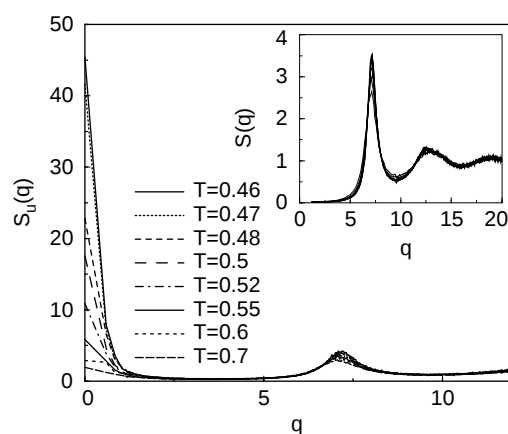


Figure 4 Displacement 'structure factor' $S_u(q, \Delta t)$ plotted versus wavevector q for different T , and for $\Delta t = \Delta t^*$. $S_u(q, \Delta t) = \langle (N\bar{\mu}^2)^{-1} \sum_{i=1}^N \sum_{j=1}^N \mu_i(t, \Delta t) \mu_j(t, \Delta t) \exp[-i\mathbf{q}(\mathbf{r}_j(t) - \mathbf{r}_i(t))] \rangle$. The values at $q = 0$ are obtained from the fluctuations in U via the relation $S_u(0, \Delta t) = \bar{\mu}^2 n T \kappa_u / \bar{\mu}^2$. Inset: static structure factor $S(q)$ for four different T , defined as $S(q) = \langle N^{-1} \sum_{i=1}^N \sum_{j=1}^N \exp[-i\mathbf{q}(\mathbf{r}_j - \mathbf{r}_i)] \rangle$.

can be detected in the particle motion. In Fig. 1, we show $g_u(r, \Delta t)$ as a function of r for $T = 0.47$ (all quantities are reported in reduced units; see legend). This temperature is within 5% of the mode-coupling critical temperature for this system, $T_c = 0.450 \pm 0.005$ (refs 19–21). The corresponding pair correlation function $g(r)$ for $T = 0.47$ is also shown. We find that for $\Delta t = 277.76$ ($\gg$ the microscopic 'collision' time), $g_u(r, \Delta t)$ is appreciably higher than $g(r)$ for values of r up to several monomer diameters. The excess correlation of the monomer displacements beyond that of their static structure is made clearer in the inset of Fig. 1, where we show the function $\Gamma(r, \Delta t) \equiv [g_u(r, \Delta t)/g(r)] - 1$.

The extent to which monomer displacements are correlated depends strongly on the choice of Δt . Figure 2a shows the total excess correlation $A(\Delta t) \equiv \int d\mathbf{r} \Gamma(r, \Delta t)$ as a function of Δt . We find that there is a time interval Δt^* at which A is a maximum and that both the maximum value of A and Δt^* increase with decreasing T . Hence for each T , the spatial correlation of monomer displacements is most prominent at Δt^* . As shown in the inset of Fig. 2a, all curves for $T \leq 0.55$ collapse onto a single master curve over several decades when t is scaled by Δt^* and A is scaled by $A(\Delta t^*)$, demonstrating that Δt^* is a characteristic time for this melt. We therefore evaluate all quantities henceforth using $\Delta t = \Delta t^*$. Figure 3a shows that Δt^* follows a power law with T : an excellent fit to the form $\Delta t^* \sim (T - T_c)^{-\gamma}$ is obtained when $T_c = 0.450 \pm 0.005$, and yields $\gamma = 1.87 \pm 0.15$. Despite the large error, this value for γ agrees better with the exponent governing the apparent vanishing of the self-diffusion coefficient D at T_c ($\gamma_D = 1.82 \pm 0.02$) than with that governing the divergence of the structural relaxation time τ_α ($\gamma_\tau = 2.09 \pm 0.07$; ref. 20).

As κ_u quantifies the total magnitude (integrated over space) of the displacement correlations measured by $G_u(\mathbf{r}, \Delta t)$, we expect κ_u to behave like A . We evaluate κ_u from equation (1) and confirm that κ_u exhibits (Fig. 2b) the same behaviour as A : κ_u goes to zero at short and long times, and has a maximum at a T -dependent Δt^* . In Fig. 3b, we show the T -dependence of A and κ_u at Δt^* . Both grow monotonically with decreasing T , indicating that the range of the correlation grows with decreasing T . We find that both quantities follow a power law with a singular temperature $T_c = 0.450$, demonstrating that $G_u(r, \Delta t^*)$ becomes long-ranged as T approaches the mode-coupling temperature.

To estimate a correlation length associated with these displacement correlations, we evaluated $S_u(q, \Delta t^*)$ for different T (Fig. 4). For intermediate and large q , $S_u(q, \Delta t^*)$ coincides with the static

structure factor $S(q)$. However, for $q \rightarrow 0$, a peak develops and grows with decreasing T , again demonstrating the presence of increasing long-range behaviour for $G_u(r, \Delta r^*)$. No such growing peak at $q = 0$ appears in $S(q)$ (Fig. 4, inset). By analogy with conventional critical phenomena, we have attempted to fit $S_u(q, \Delta r^*)$ by an Ornstein–Zernike form, $S_u(q) \propto 1/(1 + \xi^2 q^2)$, where ξ is the correlation length¹⁶. This form fits well at the highest T , but fails on approaching T_c , possibly because of finite size effects. Larger simulations may be required to determine accurately the correct functional form for S_u at small q . Nevertheless, the data show unambiguously that as $T \rightarrow T_c$, spatial correlations between the displacement of monomers become increasingly long-ranged.

Our simulations reveal a dynamical length scale relevant both to the mode-coupling dynamical transition and the glass transition. In this way we provide a bridge between the phenomenon of dynamical heterogeneity and current theories of supercooled liquids and vitrification. Remarkably, our findings are qualitatively identical to new results for simulated $\text{Ni}_{80}\text{P}_{20}$, a model metallic glassformer¹⁴, showing that the correlated monomer motion above T_c is neither due to nor strongly affected by chain connectivity. Instead, the striking similarity between $\text{Ni}_{80}\text{P}_{20}$ and the present polymer melt suggests that correlated motion is a universal feature of (at least fragile¹) glass-forming liquids. Furthermore, we have identified a fluctuating dynamical variable U in this polymer melt whose fluctuations become long-ranged and appear to diverge at T_c , and which thus behaves much like a static order parameter on approaching a second-order phase transition—albeit one that is not obviously accessible to traditional scattering experiments, but may be measurable in optical microscopy experiments on colloidal suspensions. Our findings suggest that substantial shifts in T_c could be observed by confining glass-forming liquids and melts, thereby limiting the divergence that can occur. Whether this could also explain the confinement-induced shifts of T_g observed experimentally^{22–26} needs to be investigated. Our results indicate that it may be possible to obtain further insight into the nature of supercooled, glass-forming liquids using an extension to dynamically defined quantities of the framework of ordinary critical phenomena. □

Received 15 December 1998; accepted 2 March 1999.

- Ediger, M. D., Angell, C. A. & Nagel, S. R. Supercooled liquids and glasses. *J. Phys. Chem.* **100**, 13200–13212 (1996).
- Yip, S. & Nelson, P. (eds) *Transport Theory. Stat. Phys.* **24**, 755–1270 (1995).
- Götze, W. & Sjögren, L. The mode coupling theory of structural relaxations. *Transport Theory Stat. Phys.* **24**, 801–853 (1995).
- Hiwatari, Y. & Muranaka, T. Structural heterogeneity in supercooled liquids and glasses. *J. Non-Cryst. Solids* **235–237**, 19–26 (1998).
- Perera, D. & Harrowell, P. A two-dimensional glass: microstructure and dynamics of a 2D binary mixture. *J. Non-Cryst. Solids* **235–237**, 314–319 (1998).
- Onuki, A. & Yamamoto, Y. Kinetic heterogeneities and non-linear rheology of highly supercooled liquids. *J. Non-Cryst. Solids* **235–237**, 34–40 (1998).
- Doliwa, B. & Heuer, A. Cage effect, local anisotropies, and dynamic heterogeneities at the glass transition: a computer study of hard spheres. *Phys. Rev. Lett.* **80**, 4915–4919 (1998).
- Kob, W., Donati, C., Plimpton, S. J., Poole, P. H. & Glotzer, S. C. Dynamical heterogeneities in a supercooled Lennard-Jones liquid. *Phys. Rev. Lett.* **79**, 2827–2930 (1997).
- Donati, C. *et al.* String-like clusters and cooperative motion in a model glass-forming liquid. *Phys. Rev. Lett.* **80**, 2338–2341 (1998).
- Donati, C., Glotzer, S. C., Poole, P. H., Kob, W. & Plimpton, S. J. Spatial correlations of mobility and immobility in a glass-forming Lennard-Jones liquid. *Phys. Rev. E*. (submitted).
- Böhmer, R. *et al.* Nature of the non-exponential primary relaxation in structural glass-formers probed by dynamically selective experiments. *J. Non-Cryst. Solids* **235–237**, 1–9 (1998).
- Cicerone, M. T., Blackburn, F. & Ediger, M. D. Anomalous diffusion of probe molecules in polystyrene: evidence for spatially heterogeneous segmental dynamics. *Macromolecules* **28**, 8224–8232 (1995).
- Hansen, J. P. & McDonald, I. R. *Theory of Simple Liquids* (Academic, London, 1986).
- Donati, C., Glotzer, S. C. & Poole, P. H. Growing spatial correlations of particle displacements in a simulated liquid on cooling toward the glass transition. *Phys. Rev. Lett.* (in the press).
- Glotzer, S. C., Jan, N., Lookman, T., Maclsaac, A. B. & Poole, P. H. Dynamical heterogeneity in the Ising spin glass. *Phys. Rev. E* **57**, 7350–7353 (1998).
- Stanley, H. E. *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1971).
- van Blaaderen, A. & Wiltzius, P. Real-space structure of colloidal hard-sphere glasses. *Science* **270**, 1177–1179 (1995).
- Leheny, R. L. *et al.* Structural studies of an organic liquid through the glass transition. *J. Chem. Phys.* **105**, 7783–7794 (1996).
- Bennemann, C., Paul, W., Binder, K. & Dünweg, B. Molecular-dynamics simulations of the thermal glass transition in polymer melts: α -relaxation behavior. *Phys. Rev. E* **57**, 843–851 (1998).
- Bennemann, C., Baschnagel, J. & Paul, W. Molecular-dynamics simulation of a glassy polymer melt: incoherent scattering function. *Eur. Phys. J. B* (in the press).

- Bennemann, C., Paul, W., Baschnagel, J. & Binder, K. Investigating the influence of different thermodynamic paths on the structural relaxation in a glass forming polymer melt. *J. Phys. Cond. Mat.* (in the press).
- Jerome, B. & Commandeur, J. Dynamics of glasses below the glass transition. *Nature* **386**, 589–592 (1997).
- Forrest, J. A., Dahnoki-Veress, K. & Dutcher, J. R. Interface and chain confinement effects on the glass transition temperature of thin polymer films. *Phys. Rev. E* **56**, 5705–5515 (1997).
- Wallace, W. E., van Zanten, J. H. & Wu, W. L. Influence of an impenetrable interface on a polymer glass-transition temperature. *Phys. Rev. E* **52**, R3329–R3332 (1995).
- Jackson, C. L. & McKenna, G. B. The glass transition of organic liquids confined to small pores. *J. Non-Cryst. Solids* **131–133**, 221–224 (1991).
- Arndt, M., Stannarius, R., Groothues, H. & Kremer, F. Length scale of cooperativity in the dynamic glass transition. *Phys. Rev. Lett.* **79**, 2077–2080 (1997).

Correspondence and requests for materials should be addressed to S.C.G. (e-mail: sharon.glotzer@nist.gov).

Global changes in intensity of the Earth's magnetic field during the past 800 kyr

Yohan Guyodo*† & Jean-Pierre Valet*

Institut de Physique du Globe de Paris, UMR CNRS 7577, 4, Place Jussieu, 75262 Paris Cedex 05, France

Recent advances in palaeomagnetic and dating techniques have led to increasingly precise records of the relative intensity of the Earth's past magnetic field at numerous field sites. The compilation and analysis of these records can provide important constraints on changes in global magnetic field intensity and therefore on the Earth's geodynamo itself. A previous compilation for the past 200 kyr integrated 17 marine records into a composite curve¹, with the geomagnetic origin of the signal supported by an independent analysis of ¹⁰Be production made on different cores². The persistence of long-term features in the Earth's magnetic intensity or the existence of long-term periodic changes cannot, however, be resolved in this relatively short time span. Here we present the integration of 33 records of relative palaeointensity^{3–19} into a composite curve spanning the past 800 kyr. We find that the intensity of the Earth's dipole field has experienced large-amplitude variations over this time period with pronounced intensity minima coinciding with known excursions in field direction, reflecting the emergence of non-dipole components. No stable periodicity was found in our composite record and therefore our data set does not support the hypothesis that the Earth's orbital parameters have a direct and strong influence on the geodynamo.

We constructed the database (Table 1) by selecting records obtained after appropriate normalization of magnetization intensity. All the studies have been published with the exception of a high-resolution sequence from core ODP-1021¹⁸. In addition to the data involved in the composite curve¹ referred to above ('Sint-200'), the present selection includes 16 new records relatively well distributed around the globe, 10 of which document the interval 600–800 kyr ago (Table 1). In most cases, dating was obtained by correlation to reference curves of oxygen-isotope ($\delta^{18}\text{O}$) variations or other palaeoclimate proxies such as low field susceptibility or density. The age model defined by the $\delta^{18}\text{O}$ reference curve of Bassinot *et al.*²⁰ has been applied to records older than 300 kyr; the reference curve of Martinson *et al.*²¹ was preferred for records younger than 300 kyr because of its higher resolution. An alternative approach was used for three cores (RC10-167, KS87-752 and P226) with no obvious correlation between the palaeoclimate proxies and the $\delta^{18}\text{O}$ reference curve. Following previous studies^{1,10}, these individual records of palaeointensity were correlated with a preliminary stack derived from the other well dated curves. In a second

* Present address: Department of Geology, University of Florida, Gainesville, Florida 32611, USA.

Table 1 Records involved in the construction of Sint-800

Record	Lat. (°N)	Long. (°E)	Sed. rate (cm kyr ⁻¹)	Normalization parameter	Other coherent normalization	Dating	Interval used (kyr)	Ref.
Set no. 1								
KET82-51	39	14	6.2	χ	ARM, SIRM	Corr. with KET8004 + tephra.	10–90	3
MD84-629	36	33	11.6	χ	ARM, SIRM	δ ¹⁸ O + tephra.	20–60	3
DED87-707	40	13	10.4	χ	ARM, SIRM	Corr. with KET8004 + tephra.	10–60	3
MD85-668	–1	46	6.2	ARM	χ, IRM	δ ¹⁸ O	20–140	4
MD85-669	2	47	5.3	ARM	χ, IRM	χ corr. with MD668	20–140	4
MD85-674	3	50	10.9	ARM	χ, IRM	χ corr. with MD668	20–140	4
ERDC113P	2	159	1.2	χ		δ ¹⁸ O	10–360	5, 6
NGC29	4	136	2.4	ARM	χ	χ correlated with NP5	0–190	7
NGC26	2	135	3.7	ARM	χ	χ correlated with NP5	0–120	7
NGC16	2	135	3.7	ARM	χ	χ correlated with NP5	0–190	7
NP7	1	138	2.0	ARM	χ	χ correlated with NP5	10–200	7
NP5	1	136	3.7	ARM	χ	δ ¹⁸ O	10–200	7
SU92-18	37	–27	3.5	ARM	χ, SIRM	δ ¹⁸ O	0–280	8
SU92-19	37	–27	3.2	ARM	χ, SIRM	Reflectance + χ correlated with SU92-18	0–280	8
ODP-768A	8	121	9	χ	ARM, IRM	δ ¹⁸ O + AMS radiocarb.	10–90	9
ODP-768B	8	121	10	χ	ARM, IRM	δ ¹⁸ O + AMS radiocarb.	10–130	9
ODP-769B	8	121	8.2	χ	ARM, IRM	δ ¹⁸ O + AMS radiocarb.	10–160	9
P012	58	–47	10	IRM	χ, ARM	δ ¹⁸ O + AMS radiocarb. + calib. Sint-200	10–180	10
P013	58	–48	10–30	IRM	χ, ARM	δ ¹⁸ O + AMS radiocarb. + calib. Sint-200	10–220	10
P094	50	–45	10	IRM	χ, ARM	δ ¹⁸ O + AMS radiocarb. + calib. Sint-200	0–110	10
Set no. 2								
ODP-983	60	–24	10.4	IRM	χ, ARM	δ ¹⁸ O	0–725	11
ODP-984	61	–24	12.6	IRM	χ, ARM	δ ¹⁸ O	0–450	11
KS87-752	–38	–38	2.4	ARM	χ	χ corr. with ODP940	310–800	12
ODP769	9	121	13	χ	ARM, IRM	δ ¹⁸ O	200–800	13
NP35	4	141	1	ARM	IRM	δ ¹⁸ O	130–700	14
RNDB75P	2	160	1.5	IRM	χ, ARM	δ ¹⁸ O	120–670	6
MD940	–6	61	1.3	ARM	χ, IRM	χ corr. with ODP709	100–800	15
ODP851	2	–110	1.8	ARM	χ, IRM	GRAPE, astro. calib.	30–800	16, 17
P226	3	–169	0.6	ARM	IRM	B/M reversal	40–780	14
ODP1021	40	–128	3.7	χ	ARM	χ, astro. calib.	10–800	18
NGC38	–15	–175	1	ARM	IRM	δ ¹⁸ O	10–400	14
RC10-167	33	–150	2.1	ARM		B/M reversal	0–800	19
NGC36	1	160	1	ARM	IRM	δ ¹⁸ O	0–550	14

χ, magnetic susceptibility; ARM, anhysteretic remanent magnetization; IRM, isothermal remanent magnetization.
Data available as Supplementary Information.

step, the age models were applied to the susceptibility records and subsequently refined with respect to the δ¹⁸O reference curve. This procedure increased the quality of the correlation by a factor of two. All the records have been linearly interpolated every 1 kyr before calculation of the stack.

Palaeointensity variations can be extracted from natural remanent magnetization (NRM) after normalization of the intensity by an appropriate magnetic parameter which activates the same magnetic fraction as that which carries the NRM. However, various magnetic parameters and different magnetization levels have been

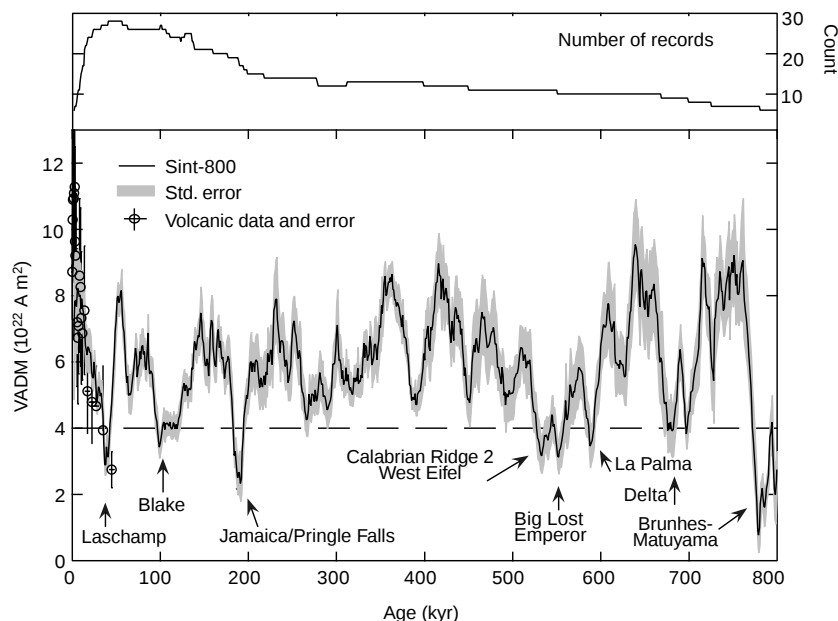


Figure 1 Synthetic record (Sint-800) with its standard error obtained from the stack of 33 records of palaeointensity. The upper part of the figure gives the number of records within successive time intervals. All the records have been dated with a common timescale and normalized (see text). The curve has been calibrated over the past 40 kyr in terms of virtual axial dipole moments (VADMs)

using averaged volcanic data (open circles) over 5-kyr intervals (correlation coefficient, $r = 0.7$). The horizontal dashed line corresponds to the critical value of intensity below which directional excursions have been observed (Langereis *et al.*²⁵). Other dips of lower amplitude coincide also with excursions reported from marine sediments.

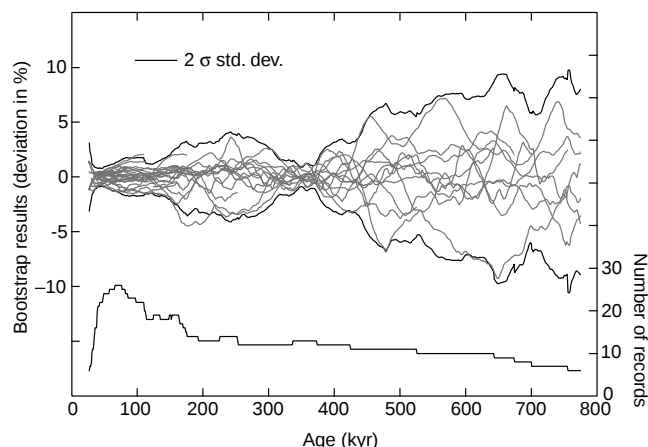


Figure 2 Tests of our synthetic record, Sint-800. Shown are results of a bootstrap process aimed at identifying drawings of 33 different composite records that show significant deviations from Sint-800. Each plot corresponds to the divergence between Sint-800 and another composite calculated after removing one record from the database. No outlier can be identified.

used by authors. Integration of the records on a common scale can therefore only be done after division of the intensities by their mean value over the entire time interval. This simple procedure accounts also for the latitudinal dependence of the dipole field intensity. As the average geomagnetic field was probably different over different periods, this normalization should be done over the same time interval for all the records. Unfortunately, no time interval is common to the entire data set which was thus split into two different groups. Records from the first group (Table 1) are younger than 350 kyr and they were normalized to unity over the interval 24–56 kyr ago. The interval 312–398 kyr ago was used for the second group. Subsequently, two independent stacks were constructed and compared over the common interval 0–350 kyr. Final integration of the two data sets was obtained after correcting the records of the first group by a factor 0.84, which corresponds to the ratio between the two averaged values.

In Fig. 1 we show our composite record ‘Sint-800’ and the associated standard error obtained after stacking all the data. In the same figure are plotted also the number of records involved over each successive time interval. The stability of the results has been tested by a bootstrap technique similar to that used for Sint-200. No significant discrepancy was observed between the stacks performed from 33 successive calculations. Deviations between Sint-800 and each successive drawing (after smoothing with a 50-kyr-long moving window to slightly reduce the resolution) lie within the limits imposed by the 2σ standard deviation (Fig. 2) and confirm the absence of significant outliers. We also note that the sharpness of the distribution is related to the number of records. Thus, some refinement (mainly a reduction of the uncertainties) would be produced by incorporation of future records covering the interval 500–800 kyr ago. Conversion into virtual axial dipole moments (VADMs) were done using volcanic records of absolute palaeointensity. Volcanic data must be averaged out over finite time intervals in order to eliminate the contribution of the non-dipole field. Only the period 0–40 kyr ago²² is sufficiently documented to allow this. According to this calibration, the highest intensities culminate at about $9 \times 10^{22} \text{ A m}^2$ and the time-averaged field value during the Brunhes chron was $(6.0 \pm 1.5) \times 10^{22} \text{ A m}^2$.

One characteristic of Sint-800 (Fig. 1) is that many intensity lows occurred during the past 800 kyr. Sint-800 was constructed from world-wide marine records with a resolution that does not exceed 3–5 kyr and is thus sensitive to the variations of the axial dipole moment. A simple condition for the occurrence of excursions^{23,24} is that the dipole field remains abnormally low so that non-dipole

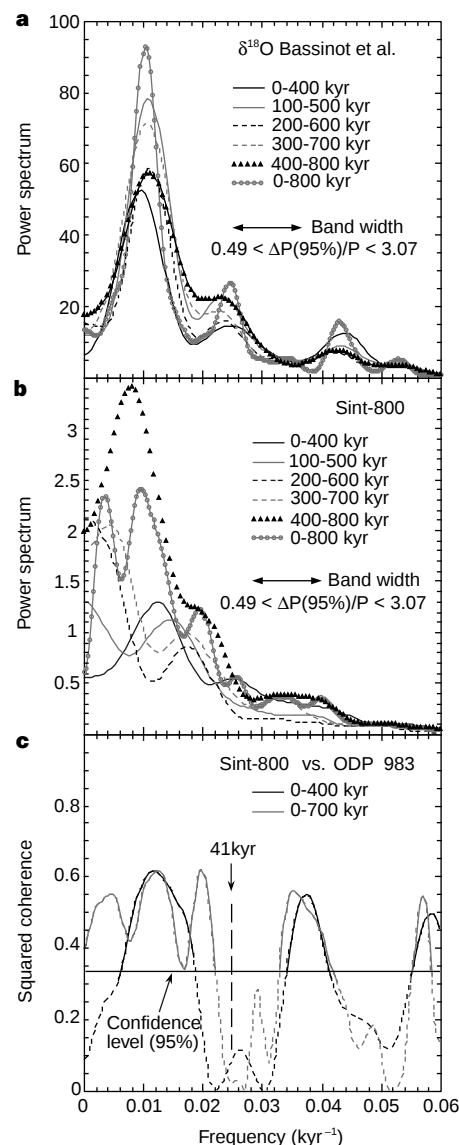


Figure 3 Spectral analysis of Sint-800 using the Blackman–Tukey method with a Bartlett window. **a**, **b**, Power spectra of the $\delta^{18}\text{O}$ curve²⁰ (**a**) and Sint-800 (**b**). Analyses have been performed over different time intervals (to explore the stability of the results). There is no stationarity or periodicity emerging from the palaeointensity variations. **c**, Coherence function between Sint-800 and ODP Site 983 (ref. 11). The calculations involve the entire 0–700 kyr interval common to the two curves and then only the 0–400 kyr interval which incorporates more records.

effects become dominant. We referred to the analysis of the short events by Langereis *et al.*²⁵ to select all the excursions that have been previously reported from both sedimentary and volcanic records. To those we added the discovery²⁶ of a large field excursion data at 595 ± 14 kyr ago from a volcanic sequence at La Palma. All these events coincide with periods of significant intensity drops (Fig. 1) and occur below a critical field of $\sim 4 \times 10^{22} \text{ A m}^2$, that is, less than 50% of the present-day field. In fact, a reduction of the present-day dipole intensity by this amount would induce the emergence of non-dipole components at many locations^{25,27}. Other excursions reported from marine sediments^{23,25,27} are synchronous with dips of lower amplitude. There is no correlation between these intensity dips and cold climate events, although such a correlation has been suggested²⁸.

The possibility that the geomagnetic field variations would be modulated by the Earth’s orbital parameters has been suggested by

several authors^{29–32}. Nevertheless, most studies dealing with long and well dated sequences conclude either that a 30–50 kyr geomagnetic pulse^{3,6} exists, or that there is no dominant period. But a correlation has been found¹¹ between the 41-kyr obliquity cycle and palaeointensity variations; this was based on the analysis of a very detailed and precisely dated sequence¹¹ from Ocean Drilling Project Site 983. We performed a spectral analysis of Sint-800 using the Blackman–Tukey technique with the AnalySeries software³³. In order to investigate the stability of the results, we first analysed the entire signal and then restrained the analysis to the 0–400 kyr interval which incorporates more records. The spectral contents of these two intervals are significantly different. The same conclusion is reached by comparing every 400-kyr-long interval in increments of 100 kyr with the 0–800 kyr period (Fig. 3). These results point out the absence of any dominant stable periodicity. For comparison, the same analysis performed with the $\delta^{18}\text{O}$ curve²⁰ shows perfect reproducibility of the orbital peaks (23, 41 and 100 kyr) over each interval. (An additional indication supporting these conclusions is that the artificial 41-kyr signal created after band-pass-filtering Sint-800 has a different phase and a much smaller amplitude than the signal obtained by the same procedure for site 983.) Similar observations and conclusions are reached without incorporating the site 983 record in the database. Orbital modulation of the geomagnetic field should primarily affect the dipole field, and therefore should be present in Sint-800 which incorporates at least 20 data points per 41 kyr. However, for periods shorter than 40 kyr it is possible that uncertainties due to dating and/or smoothing inherent to the stacking process induced loss of spectral power (Fig. 3b). This may explain the pattern observed in the coherence functions between Sint-800 and data from site 983 (Fig. 3c) for these short periods.

Our composite Sint-800 curve shows that the Earth's dipole field over the past 800 kyr was dominated by changes of very large (and various) amplitude but does not indicate the presence of any dominant periodicity. The mean field value remained more or less constant. Geomagnetic excursions are observed when the dipole moment decreases to a critical value of about $4 \times 10^{22} \text{ A m}^2$, and such excursions must thus be seen as direct consequences of the overall 'secular' variation of the dipole field. □

Received 11 January; accepted 29 March 1999.

- Guyodo, Y. & Valet, J.-P. Relative variations in geomagnetic intensity from sedimentary records: the past 200,000 years. *Earth Planet. Sci. Lett.* **143**, 23–36 (1996).
- Frank, M. *et al.* A 200 kyr record of cosmogenic radionuclide production rate and geomagnetic field intensity from ^{10}Be in globally stacked deep-sea sediments. *Earth Planet. Sci. Lett.* **149**, 121–129 (1997).
- Tric, E. *et al.* Paleointensity of the geomagnetic field for the last 80,000 years. *J. Geophys. Res.* **97**, 9337–9351 (1992).
- Meynadier, L., Valet, J.-P., Weeks, R., Shackleton, N. & Hage, V. L. Relative geomagnetic intensity of the field during the last 140 ka. *Earth Planet. Sci. Lett.* **114**, 39–57 (1992).
- Tauxe, L. & Wu, G. Normalised remanence in sediments from western equatorial Pacific: relative paleointensity of the geomagnetic field. *J. Geophys. Res.* **95**, 12337–12350 (1990).
- Tauxe, L. & Shackleton, N. J. Relative paleointensity records from the Ontong-Java Plateau. *Geophys. J. Int.* **117**, 769–782 (1994).
- Yamazaki, T. & Ioka, N. Long-term secular variation of the geomagnetic field during the last 200 k.y. recorded in sediment cores from the western equatorial Pacific. *Earth Planet. Sci. Lett.* **128**, 527–544 (1994).
- Lehman, B. *et al.* Relative changes of the geomagnetic field intensity during the last 280 kyr from piston cores in the Acores area. *Phys. Earth Planet. Inter.* **93**, 269–284 (1996).
- Schneider, D. A. Late Pleistocene and Holocene geomagnetic intensity variations recorded in Sulu Sea sediments. *Eos (suppl.)* **75**, S119 (1994).
- Stoner, J. S., Channell, J. E. T. & Hillaire-Marcel, C. A 200 ka chronostratigraphy for the Labrador Sea: Indirect correlation of the sediment record to SPECMAP. *Earth Planet. Sci. Lett.* **159**, 165–181 (1998).
- Channell, J. E. T., Hodell, D. A., McManus, J. & Lehman, B. Orbital modulation of the Earth's magnetic field intensity. *Nature* **394**, 464–468 (1998).
- Valet, J.-P., Meynadier, L., Bassinot, F. & Garnier, F. Relative paleointensity across the last geomagnetic reversal from sediments of the Atlantic, Indian and Pacific Oceans. *Geophys. Res. Lett.* **21**, 485–488 (1994).
- Schneider, D. A. & Mello, G. A. A high-resolution marine sedimentary record of geomagnetic intensity during the Brunhes Chron. *Earth Planet. Sci. Lett.* **144**, 297–314 (1996).
- Yamazaki, T., Ioka, N. & Eguchi, N. Relative paleointensity of the geomagnetic field during the Brunhes Chron. *Earth Planet. Sci. Lett.* **136**, 525–540 (1995).
- Meynadier, L., Valet, J.-P., Bassinot, F., Shackleton, N. & Guyodo, Y. Asymmetrical saw-tooth pattern of the geomagnetic field intensity from equatorial sediments in the Pacific and Indian Oceans. *Earth Planet. Sci. Lett.* **126**, 109–127 (1994).
- Meynadier, L. & Valet, J.-P. Relative geomagnetic intensity during the last 4 m.y. from the equatorial Pacific. *Proc. ODP Sci. Res.* **138**, 779–795 (1995).

- Meynadier, L., Valet, J.-P., Guyodo, Y. & Richter, C. Saw-toothed variations of relative paleointensity and cumulative viscous remanence: Testing the records and the model. *J. Geophys. Res.* **103**, 7095–7105 (1998).
- Guyodo, Y., Richter, C. & Valet, J.-P. Paleointensity record from the Pleistocene sediments (1.4–0 Ma) off the California Margin. *J. Geophys. Res.* (in the press).
- Kent, D. V. & Opdyke, N. D. Paleomagnetic field intensity recorded in a Brunhes epoch deep-sea sediment core. *Nature* **266**, 156–159 (1977).
- Bassinot, F. C. *et al.* The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth Planet. Sci. Lett.* **126**, 91–108 (1994).
- Martinson, D. G. *et al.* Age dating and the orbital theory of the Ice Ages: development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).
- Merrill, R. T., McElhinny, M. W. & McFadden, P. L. (eds) *The Magnetic Field of the Earth* (Int. Geophysics Ser., Academic, London, 1998).
- Valet, J.-P. & Meynadier, L. Geomagnetic field intensity and reversals during the past four million years. *Nature* **366**, 91–95 (1993).
- Merrill, R. T. & McFadden, P. L. Geomagnetic stability: Reversal events and excursions. *Earth Planet. Sci. Lett.* **121**, 57–69 (1994).
- Langeris, C. G., Dekkers, M. J., de Lange, G. J., Paterne, M. & van Santvoort, P. J. M. Magnetotratigraphy and astronomical calibration of the last 1.1 Myr from an eastern Mediterranean piston core and dating of short events in the Brunhes. *Geophys. J. Int.* **129**, 75–94 (1997).
- Quidelleur, X., Gillot, P. Y. & Carlot, J. Link between excursions and paleointensity inferred from abnormal field directions recorded at La Palma around 600 ka. (suppl.) **78**, F181 (1997).
- Champion, D. E., Lanphere, M. A. & Kuntz, M. A. Evidence for a new geomagnetic reversal from lava flows in Idaho: Discussion of short polarity reversals in the Brunhes and late Matuyama polarity chrons. *J. Geophys. Res.* **93**, 11667–11680 (1988).
- Worm, H.-U. A link between geomagnetic reversals, events and glaciations. *Earth Planet. Sci. Lett.* **177**, 55–67 (1997).
- Malkus, W. V. R. Precession of the earth as the cause of geomagnetism. *Science* **160**, 259–264 (1968).
- Wollin, G., Ericson, D. B. & Ryan, W. B. F. Magnetism of the earth and climatic changes. *Earth Planet. Sci. Lett.* **12**, 171–183 (1971).
- Creer, K. M., Thouveny, N. & Blunk, I. Climatic and geomagnetic influences on the Lac du Bouchet paleomagnetic record through the last 110,000 years. *Phys. Earth Planet. Inter.* **64**, 314–341 (1990).
- Vanyo, J. P. A geodynamo powered by luni-solar precession. *Geophys. Astrophys. Fluid Dyn.* **59**, 209–234 (1991).
- Paillard, D., Labeyrie, L. & Yiou, P. Macintosh program performs time-series analysis. *Eos* **77**, 379 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank J. E. T. Channell, C. Laj, J. Meynadier, D. A. Schneider, J. S. Stoner, L. Tauxe and T. Yamazaki for supplying data, and F. C. Bennett for comments. This work was supported by the CNRS-INSU program Interieur de la Terre. At this University of Florida, Y.G. was supported by a grant awarded to J. E. T. Channell.

Correspondence and requests for material should be addressed to Y.G. (e-mail: guyodo@ufl.edu). Data from sites 983 and 984 are available from J.E.T. Channell at the University of Florida (e-mail: jetc@nerdc.ufl.edu).

Nature of the Earth's earliest crust from hafnium isotopes in single detrital zircons

Yuri Amelin*, Der-Chuen Lee†, Alex N. Halliday‡ & Robert T. Pidgeon§

* Department of Earth Sciences, Royal Ontario Museum, Toronto, Ontario M5S 2C6, Canada

† Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA

‡ Department of Earth Sciences, ETH, Zurich, CH-8092, Switzerland

§ School of Applied Geology, Curtin University of Technology, Bentley, WA 6102, Australia

Continental crust forms from, and thus chemically depletes, the Earth's mantle. Evidence that the Earth's mantle was already chemically depleted by melting before the formation of today's oldest surviving crust has been presented in the form of Sm–Nd isotope studies of 3.8–4.0 billion years old rocks from Greenland^{1–5} and Canada^{5–7}. But this interpretation has been questioned because of the possibility that subsequent perturbations may have re-equilibrated the neodymium-isotope compositions of these rocks⁸. Independent and more robust evidence for the origin of the earliest crust and depletion of the Archaean mantle can potentially be provided by hafnium-isotope compositions of zircon, a mineral whose age can be precisely determined by U–Pb dating, and which can survive metamorphisms⁴. But the amounts of hafnium in single zircon grains are too small for the isotopic composition to be precisely analysed by conventional methods.

Here we report hafnium-isotope data, obtained using the new technique of multiple-collector plasma-source mass spectrometry⁹, for 37 individual grains of the oldest known terrestrial zircons (from the Narryer Gneiss Complex, Australia, with U–Pb ages of up to 4.14 Gyr (refs 10–13)). We find that none of the grains has a depleted mantle signature, but that many were derived from a source with a hafnium-isotope composition similar to that of chondritic meteorites. Furthermore, more than half of the analysed grains seem to have formed by remelting of significantly older crust, indicating that crustal preservation and subsequent reworking might have been important processes from earliest times.

Zircon is well suited for Hf-isotope studies of ancient rocks. It has high Hf concentrations and low Lu/Hf ratios, so that correction for *in situ* radiogenic growth is negligible. It can be precisely dated by U–Pb techniques and these data can confirm whether the zircon remained a closed system. Zircon is extremely resistant and can survive erosion or metamorphism that might modify or destroy its host rock. Therefore, zircon has been successfully used in many Lu–Hf isotope studies^{4,14,15}. But at least 1–2 μg of Hf (and therefore 100–200 μg of zircon) is normally required for precise analysis. This precludes studies of detrital zircons where grains come from various sources and must be analysed individually. The recently developed multiple-collector inductively coupled plasma-source mass spectrometer (MC-ICPMS) allows precise isotopic analysis of 30–50 ng Hf (refs 16,17), the amount of Hf present in a single 3–5- μg zircon grain. This level of analytical sensitivity allows us, we believe for the first time, to analyse Hf grain by grain from a detrital zircon population rather than in bulk¹⁸, and to elucidate the earliest history of the continental crust.

Newly formed crust inherits the Hf-isotope composition of its mantle source, and this can be relatively unradiogenic ($\epsilon_{\text{Hf}}(T) < 0$) or radiogenic ($\epsilon_{\text{Hf}}(T) > 0$) Hf depending on whether the mantle from which it was derived was respectively enriched or depleted ($\epsilon_{\text{Hf}}(T)$ is defined in Fig. 2 legend). Zircons crystallized from felsic magmas formed by melting of continental crust have initial Hf-isotope compositions which indicate the Hf-isotope composition of a particular sample of the continental crust. For example, if granites were formed by melting of juvenile crust that had recently formed from depleted mantle, as in many modern arcs, the zircons would have radiogenic initial Hf-isotope compositions ($\epsilon_{\text{Hf}}(T) > 0$) close to that of the mantle source. However, if granite formation occurred by reworking of old continental crust, the newly crystallized zircon

Hf signature would be unradiogenic ($\epsilon_{\text{Hf}}(T) < 0$).

At present, nearly all production of continental crust takes place by melting of mantle with a time-integrated history of depletion. This depletion is itself caused by extraction of crust from the mantle, so it provides an indirect monitor of the history of continental growth. Determining the record of mantle depletion is therefore of great importance. To compare recent crust–mantle differentiation with that of the early Earth we have investigated the Hf-isotope signatures of the oldest known detrital zircons from the Archaean Narryer Gneiss Complex in western Australia.

This complex includes two localities of sedimentary rocks containing unusually old zircons: the Mount Narryer quartzite¹⁰ and the Jack Hill metaconglomerate¹¹. The latter is particularly significant because the fraction of zircon grains older than 4 Gyr is relatively large: 12–20% (refs 1, 13), some being as old as 4.28 Gyr (ref. 11). The trace-element concentrations and the mineralogy of inclusions in the Jack Hills detrital zircons, both ‘old’ and ‘young’, confirm the view that these zircons grew in felsic magmas¹². Although the Jack Hills zircons have been dated by U–Pb and analysed for trace-element concentrations¹², no Lu–Hf-isotope determinations have been reported so far. An attempt to determine the Hf-isotope compositions of similar 4.20–3.75-Gyr Mount Narryer zircons using ion microprobe techniques¹⁹ yielded data of low precision ($\pm(5-7)\epsilon$ units at 2σ), insufficient to resolve enriched or depleted crust and mantle components.

We have studied 37 single detrital zircon grains from the Jack Hills metaconglomerate for Lu–Hf and U–Pb (see table in Supplementary Information). The U–Pb ages range up to 4.14 Gyr (Fig. 1). Almost all of our U–Pb analyses are less than 10% discordant (Fig. 1 and Supplementary Information) and about half are less than 3% discordant. This is important because in a detrital single-grain study, where each grain is potentially unique, the reliability of $\epsilon_{\text{Hf}}(T)$ cannot be checked by analyses of zircon fractions with varying U–Pb discordance and by abrasion treatments, as has become standard practice in multigrain isotope studies^{4,15}. The

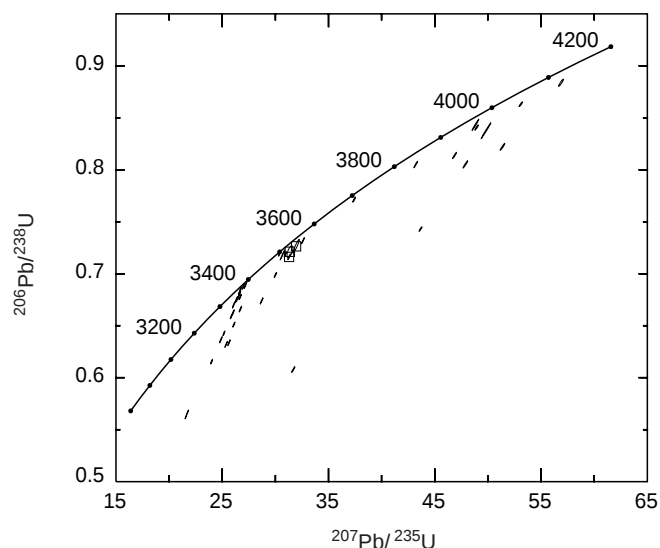


Figure 1 U–Pb concordia diagram for the single zircon grains and fragments analysed in this study. Error ellipses are 2σ . Three data points from Acasta gneiss zircons are marked with boxes. Numbers 3200–4200 are concordia age labels, millions of years.

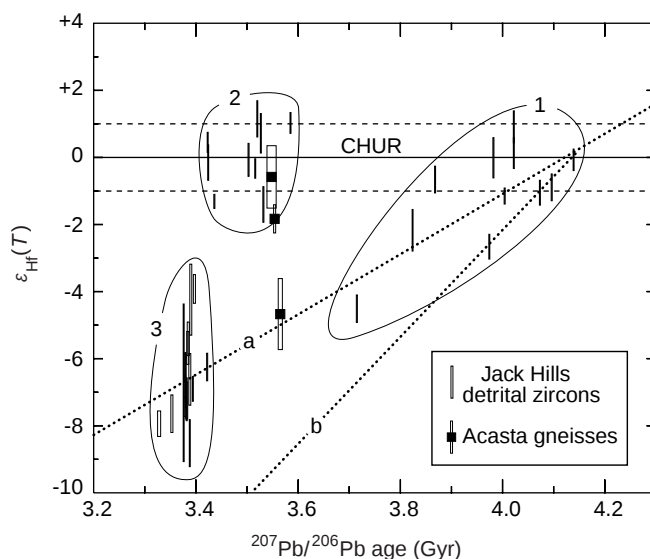


Figure 2 Hf-isotope evolution diagram for the Jack Hills and Acasta zircons. Error bars are 2σ . The Acasta data are marked with boxes. The chondritic reference line (CHUR) and its error limits (dashed lines) are from ref. 16. Groups of the Jack Hills data points (1, 2 and 3) are discussed in the text. Dotted line a is the best-fit line through the Jack Hills data points of pre-3.7-Gyr and ~3.38-Gyr groups, corresponding to the evolution of mafic crust with $^{176}\text{Lu}/^{177}\text{Hf} = 0.022$ (see text). Dotted line b is the evolution trend of average granitoid crust with $^{176}\text{Lu}/^{177}\text{Hf} = 0.0093$ (ref. 27). $\epsilon_{\text{Hf}}(T)$ is a deviation, multiplied by 10^4 , of initial Hf isotopic ratio from the chondritic reference value (ref. 16).

Lu–Hf system in zircon is robust, and the only potential complication for initial $^{176}\text{Lu}/^{177}\text{Hf}$ determination would be the presence of domains with different ages and Hf-isotope compositions, for example, cores and overgrowths^{14,20}. This problem is minimized by analysing zircons with little or no U–Pb discordance^{4,14,15,20}.

Accurate determination of $\epsilon_{\text{Hf}}(T)$ can be compromised by using a biased age value even for a zircon with uniform Hf-isotope composition, and this effect should also be evaluated. The effect of age on calculated $\epsilon_{\text{Hf}}(T)$ is 2.2–2.5 ϵ units per 100 Myr, due to the age dependence of the chondritic reference value. This problem may affect zircon that experienced ancient Pb loss. In such a case, the $^{207}\text{Pb}/^{206}\text{Pb}$ age is lower than the crystallization age, while the Hf-isotope composition is not modified, so the apparent $\epsilon_{\text{Hf}}(T)$ is shifted towards more negative values. Previous U–Pb data^{11–13} have shown internal age heterogeneity in the Jack Hills zircon grains of the ‘old’ (>3.8 Gyr) group. The possible bias in $\epsilon_{\text{Hf}}(T)$ due to ancient Pb loss can be estimated from the difference between the ‘whole grain’ $^{207}\text{Pb}/^{206}\text{Pb}$ age reconstructed from fragment analyses, and the age of the oldest fragment taken as a proxy for crystallization age. Age biases between ‘whole grains’ and the oldest fragments for the five fragmented 3.8–4.1-Gyr Jack Hills zircons¹³ are between 8 and 61 Myr, corresponding to shifts in $\epsilon_{\text{Hf}}(T)$ between 0.2 and 1.5 ϵ units. The multiple spot ion microprobe analyses of ‘old’ Jack Hills zircons^{11,12} indicate similar or smaller internal age variations, so a bias of 1.5 ϵ units is a maximum estimate, and is likely to be much smaller for most grains.

The hafnium-isotope data are plotted as a function of age in Fig. 2. The data fall into three groups (Fig. 2): (1) 3.72–4.14-Gyr grains with $\epsilon_{\text{Hf}}(T)$ between +0.9 and –4.5; (2) 3.42–3.58-Gyr grains with $\epsilon_{\text{Hf}}(T)$ of +1.1 to –1.4; and (3) ~3.38-Gyr grains with $\epsilon_{\text{Hf}}(T)$ between –3.9 and –8.5. Five grains of the first group with $^{207}\text{Pb}/^{206}\text{Pb}$ ages >4.0 Gyr have nearly chondritic $\epsilon_{\text{Hf}}(T)$, from +0.9 to –1.1. The 3.7–4.0-Gyr grains have slightly lower $\epsilon_{\text{Hf}}(T)$ values that correlate with age. The 3.38-Gyr group plots on the extension of this trend. Taken together, the 3.38-Gyr and pre-3.7-Gyr grains form a scattered array in a plot of $\epsilon_{\text{Hf}}(T)$ versus apparent age. The data for ~3.6-Gyr zircons plot as a compact group well above this array: their range of $\epsilon_{\text{Hf}}(T)$ is relatively limited (2.5 ϵ units) and centres on the chondritic reference line.

Using these data, we can address two questions: the extent of depletion of the terrestrial mantle before 4 Gyr, and the nature of the earliest Archaean continental crust. It is clear that none of the 37 grains analysed appear to have come from a source that carried a significant record of depletion of the Earth’s mantle. That is, no zircons have strongly positive $\epsilon_{\text{Hf}}(T)$. The complete absence of depleted mantle signatures is in marked contrast with uniformly positive $\epsilon_{\text{Hf}}(T)$ values in the Itsaq Gneiss Complex of west Greenland^{4,21}, the only early Archaean suite of rocks studied for Hf-isotope systematics.

The second notable feature of the data is the large proportion of zircons (more than half) that are derived from sources with time-integrated enrichment (negative $\epsilon_{\text{Hf}}(T)$). Clearly, much of the crust was already of a significant age when these zircons grew. The $\epsilon_{\text{Hf}}(T)$ of magmas derived by episodic remelting of a particular crustal section would decrease with time because the $^{176}\text{Lu}/^{177}\text{Hf}$ ratio of the crust is lower than the chondritic value. It has been proposed that the earliest crust was mafic^{22,23}. The $^{176}\text{Lu}/^{177}\text{Hf}$ ratios in crustal mafic rocks, including Archaean basalts and komatiites, are mostly between 0.015 and 0.028 (refs 24–26), while $^{176}\text{Lu}/^{177}\text{Hf}$ ratios in Precambrian granitoids are lower still, between 0.001 and 0.023, and averaging 0.0093 (based on the data from ref. 27). The $\epsilon_{\text{Hf}}(T)$ versus time array of the 3.38-Gyr and pre-3.7-Gyr zircons (Fig. 2) can be interpreted to reflect the evolution of a crustal block, derived from the mantle at or before 4.14 Gyr. The slope of this array corresponds to $^{176}\text{Lu}/^{177}\text{Hf} = 0.022$, consistent with the above range for crustal mafic rocks. It appears that very old (>4.14–4.02-Gyr) crust underwent a protracted history of remelting at 4.0–3.97, 3.8–3.7

and 3.38 Gyr. Complex internal age patterns of the ‘old’ zircon group^{11,13} indicate involvement of earlier granitoids in later melting episodes within this block. The 3.38-Gyr episode, which produced the most abundant fraction of the Jack Hills zircon population, was probably a major orogeny. The range of zircon ages indicates a duration of 20–30 Myr, and the range of initial Hf compositions of over 4.5 ϵ units implies extensive mixing between pre-4.0 Gyr crust and younger crust.

The unradiogenic Hf-isotope compositions of some of the ~4-Gyr zircon grains allow evaluation of a minimum crustal residence time for the protoliths of parent rocks of the zircons. The $\epsilon_{\text{Hf}}(T)$ value of –2.7 for the 3.97-Gyr grain no. 40 requires a crustal residence time of 150 Myr before the zircon crystallization, if a crustal block with average granitoid $^{176}\text{Lu}/^{177}\text{Hf} = 0.0093$ (ref. 27) was derived from mantle with chondritic⁵ Lu–Hf. This minimum estimate is essentially model-independent. Recently suggested alternative chondritic reference parameters^{28,29}, or assumption of a depleted mantle source, or higher $^{176}\text{Lu}/^{177}\text{Hf}$ in the crustal block (more mafic crust) would all result in a larger value for the crustal residence time. Thus not only zircon grains but also crustal blocks, formed as early as 4.12–4.14 Gyr, were preserved for 150 Myr or longer, rather than being recycled immediately to the mantle. Remnants of these crustal blocks, formed ~100 Myr before the oldest terrestrial crustal rocks known^{30,31}, may still be found.

All data points of pre-4.0-Gyr zircons, as well as most of the ~3.5-Gyr group, plot at or very close to the chondritic reference line (the range of $\epsilon_{\text{Hf}}(T)$ from +1 to –1). Although granitoids with $\epsilon_{\text{Hf}}(T) \approx 0$ might be produced by mixing between older depleted mantle and enriched crust derivatives, such mixing would probably result in a broad range of initial isotope ratios from highly positive to strongly negative, contrasting with the consistent upper limit of $\epsilon_{\text{Hf}}(T) \approx 0$ in the Jack Hills zircons. It is also possible that primary crust was derived from depleted mantle, but that crustal melting was fortuitously retarded for 200–300 Myr or more, allowing $\epsilon_{\text{Hf}}(T)$ to evolve from positive to nearly zero values, but this *ad hoc* explanation is not predicted by comparisons with modern arcs. Therefore, we consider the most straightforward interpretation of the data as reflecting the derivation of primary crust from a mantle reservoir with nearly chondritic Hf-isotope composition, followed by early crustal differentiation. A chondritic Hf-isotope composition does not mean that the mantle was never melted. It could, for example, reflect an efficient large-scale mixing between depleted and enriched mantle. Our results are compatible with a chondritic mantle model and provide no indication that the bulk silicate earth has been fractionated in Lu/Hf by perovskite fractionation in a magma ocean²⁸, such that a chondritic reference model for the source of basalts might be inaccurate.

Our data set for 37 early Archaean detrital zircons indicates the need for further comparative studies. We have already analysed three ~3.55-Gyr zircon grains from two samples of Acasta gneisses (northwest Canada; see Supplementary Information). The $\epsilon_{\text{Hf}}(T)$ values are heterogeneous and vary from –0.6 to –4.7, also indicating a long crustal pre-history, consistent with previous geochronological and Nd, Pb isotope data^{6,7,30,31}. The similarities in Hf-isotope composition between the Acasta gneisses and some of the source rocks of the Jack Hills conglomerate are consistent with previous suggestions of a relationship, based on U–Pb data alone^{13,30,31}. Again, these preliminary data yield no indication of a depleted mantle component.

Depleted mantle and enriched mafic or felsic crust should have been produced as complementary reservoirs in the early Archaean, so finding isotopic signals from both enriched and depleted reservoirs can be expected. Current evidence for early Archaean depleted mantle, based upon Hf-isotope data from the Itsaq Gneiss Complex, has been presented by Vervoort *et al.*⁴ and by Vervoort and Blichert-Toft²¹. However, the zircon populations of these polymetamorphic rocks are very complex, and obtaining the most reliable Hf-isotope

data would require integrated Lu–Hf and U–Pb study of single grains or even fragments of grains. The conclusion of these authors²¹ that “there are no verified cases where any pre-3.5 Ga rocks have been derived from an older enriched reservoir” is inconsistent with our data which provide unequivocal evidence that such a reservoir did exist, and persisted for hundreds of Myr.

The results of our study raise questions concerning whether early Archaean depleted mantle was global or regional in extent, whether it formed and was rapidly remixed with less depleted components by convection, or whether it became inaccessible as a subsequent source of new crust. These are fundamental issues for future single-zircon Lu–Hf research. Our data demonstrate that detrital zircon populations contain a wealth of information pertinent to these issues, which can be most reliably recovered using a combination of precise U–Pb and Lu–Hf isotope measurements. Some previous interpretations of the extent of early mantle depletion based on whole-rock studies, bulk zircon populations and less-precise analytical methods may need re-evaluation.

Methods

All zircons were hand-picked to avoid inclusions, fractures or turbidity, and air-abraded³². Some grains were broken into fragments¹³. Zircons were spiked with mixed ²³⁵U–²⁰⁵Pb and digested in HF–HNO₃ in sealed bombs³³ for 2–4 d at 200 °C. Samples were dried and redissolved in 3.1 M HCl. Separation of U and Pb was done on AG1x8 (200–400 mesh; Bio-Rad, Richmond, California) in HCl and H₂O (refs 13, 33). Spiking with mixed ¹⁷⁶Lu–¹⁸⁰Hf (and sometimes ¹⁴⁹Sm–¹⁵⁰Nd) was done at various stages: after U–Pb chemical separation (grains 1–12), before digestion (grains 40–53 and zircons from Acasta gneiss), or on 95% aliquots in 3.1 M HCl before U–Pb separation (grains 54–65). Experiments with two standard zircon samples (Y.A. *et al.*, manuscript in preparation) have shown that all three spiking methods yield identical Lu/Hf ratios. After spiking for Lu–Hf, all samples were returned to the 200 °C oven with 1.0 M HCl–0.1 M HF to ensure sample-spike equilibration. Two-column separation of Lu and Hf follows the procedure described in ref. 15, except for the use of a smaller size of cation-exchange columns and the use of Ln-Spec resin (Eichrom, Dalriem, Illinois) instead of Teflon covered by HDEHP (bis-(2-ethylhexyl) hydrogen phosphate) to separate Hf from Zr. U, Pb and Lu were analysed at the Royal Ontario Museum by thermal ionization mass spectrometry (TIMS) using mostly an analogue Daly photomultiplier detector^{13,15}. Hf was analysed by MC-ICPMS in static multicollector mode at the University of Michigan^{9,34}. The JMC-475 standard was run between samples to monitor instrument performance and ensure negligible memory. Hf-isotope ratios of spiked samples were calculated from analyses by off-line numeric solution of isotope dilution equations with exponential normalization to ¹⁷⁹Hf/¹⁷⁷Hf = 0.7325. Procedure blanks of 0.2–0.5 pg Lu and 20–30 pg Hf were negligible for all samples.

Received 24 August 1998; accepted 19 March 1999.

- Hamilton, P. J., O’Nions, R. K., Evensen, N. M., Bridgwater, D. & Allaart, H. Sm–Nd isotopic investigations of Isua supracrustals and implications for mantle evolution. *Nature* **272**, 41–43 (1978).
- Jacobsen, S. B. & Dymek, R. F. Nd and Sr isotope systematics of clastic metasediments from Isua, West Greenland: identification of pre-3.8 Ga differentiated crustal components. *J. Geophys. Res.* **93**, 338–354 (1988).
- Bennett, V. C., Nutman, A. P. & McCulloch, M. T. Nd isotopic evidence for transient, highly depleted mantle reservoirs in the early history of the Earth. *Earth Planet. Sci. Lett.* **119**, 299–317 (1993).
- Vervoort, J. D., Patchett, P. J., Gehrels, G. E. & Nutman, A. P. Constraints on the early Earth differentiation from hafnium and neodymium isotopes. *Nature* **379**, 624–627 (1996).
- Moorbath, S., Whitehouse, M. J. & Kamber, B. S. Extreme Nd-isotope heterogeneity in the early Archaean—fact or fiction? Case histories from northern Canada and West Greenland. *Chem. Geol.* **135**, 213–231 (1997).
- Browning, S. A., King, J. E., Housh, T. B., Isachsen, C. E. & Podosek, F. A. Neodymium and lead isotope evidence for enriched early Archaean crust in North America. *Nature* **340**, 222–225 (1989).
- Browning, S. A. & Housh, T. B. The Earth’s early evolution. *Science* **269**, 1535–1540 (1995).
- Gruau, G., Rosing, M., Bridgwater, D. & Gill, R. C. O. Resetting of Sm–Nd systematics during metamorphism of >3.7-Ga rocks: implications for isotopic models of early Earth differentiation. *Chem. Geol.* **133**, 225–240 (1996).
- Halliday, A. N. *et al.* Recent developments in inductively coupled plasma magnetic sector multiple collector mass spectrometry. *Int. J. Mass Spec. Ion Proc.* **146/147**, 21–33 (1995).
- Froude, D. O., Ireland, T. R., Kinny, P. D., Williams, I. S. & Compston, W. Ion microprobe identification of 4,100–4,200 Myr-old terrestrial zircons. *Nature* **304**, 616–618 (1983).
- Compston, W. & Pidgeon, R. T. Jack Hills, evidence of more very old detrital zircons in Western Australia. *Nature* **321**, 766–769 (1986).
- Maas, R., Kinny, P. D., Williams, I. S., Froude, D. O. & Compston, W. The Earth’s oldest known crust: A geochronological and geochemical study of 3900–4200 Ma old detrital zircons from Mt. Narryer and Jack Hills, Western Australia. *Geochim. Cosmochim. Acta* **56**, 1281–1300 (1992).

- Amelin, Y. V. Geochronology of the Jack Hills detrital zircons by precise U–Pb isotope dilution analysis of crystal fragments. *Chem. Geol.* **146**, 25–38 (1998).
- Patchett, P. J., Kouvo, O., Hedge, C. E. & Tatsumoto, M. Evolution of continental crust and mantle heterogeneity: evidence from Hf isotopes. *Contrib. Mineral. Petrol.* **78**, 279–297 (1981).
- Corfu, F. & Noble, S. R. Genesis of southern Abitibi greenstone belt, Superior Province, Canada: Evidence from zircon Hf isotope analyses using a single filament technique. *Geochim. Cosmochim. Acta* **56**, 2081–2097 (1992).
- Blichert-Toft, J. & Albarède, F. The Lu–Hf isotope geochemistry of chondrites and the evolution of the crust–mantle system. *Earth Planet. Sci. Lett.* **148**, 243–258 (1997).
- Blichert-Toft, J., Chauvel, C. & Albarède, F. Separation of Hf and Lu for high-precision isotope analysis of rock samples by magnetic sector–multiple collector ICP–MS. *Contrib. Mineral. Petrol.* **127**, 248–260 (1997).
- Stevenson, R. K. & Patchett, P. J. Implications for the evolution of continental crust from Hf isotope systematics of Archaean detrital zircons. *Geochim. Cosmochim. Acta* **54**, 1683–1697 (1990).
- Kinny, P. D., Compston, W. & Williams, I. S. A reconnaissance ion–probe study of hafnium isotopes in zircons. *Geochim. Cosmochim. Acta* **55**, 849–859 (1991).
- Patchett, P. J. Importance of Lu–Hf isotopic system in studies of planetary chronology and chemical evolution. *Geochim. Cosmochim. Acta* **47**, 81–91 (1983).
- Vervoort, J. D. & Blichert-Toft, J. Evolution of the depleted mantle: Hf isotope evidence from juvenile rocks through time. *Geochim. Cosmochim. Acta* (in the press).
- Chase, C. G. & Patchett, P. J. Stored mafic/ultramafic crust and early Archaean mantle depletion. *Earth Planet. Sci. Lett.* **91**, 66–72 (1988).
- Galer, S. J. G. & Goldstein, S. L. Early mantle differentiation and its thermal constraints. *Geochim. Cosmochim. Acta* **55**, 227–239 (1991).
- Taylor, S. R. & McLennan, S. M. *The Continental Crust, its Composition and Evolution* (Blackwell Scientific, Oxford, 1985).
- Condie, K. C. Chemical composition and evolution of the upper continental crust: contrasting results from surface samples and shales. *Chem. Geol.* **104**, 1–37 (1993).
- Rudnick, R. L. & Fountain, D. M. Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
- Vervoort, J. D. & Patchett, P. J. Behaviour of hafnium and neodymium isotopes in the crust: constraints from Precambrian crustally derived granites. *Geochim. Cosmochim. Acta* **60**, 3717–3733 (1996).
- Salter, V. J. M. & White, W. M. Hf isotopic constraints on mantle evolution. *Chem. Geol.* **145**, 447–460 (1998).
- Vervoort, J. D., Patchett, P. J., Blichert-Toft, J. & Albarède, F. Relationships between Lu–Hf and Sm–Nd isotopic systems in the global sedimentary system. *Earth Planet. Sci. Lett.* (in the press).
- Stern, R. A. & Bleeker, W. Age of the world’s oldest rocks refined using Canada’s SHRIMP: the Acasta Gneiss Complex, Northwest Territories, Canada. *Geosci. Canada* **25**, 27–31 (1998).
- Bowring, S. S. & Williams, I. S. Priscoan (4.00–4.03 Ga) orthogneisses from northwestern Canada. *Contrib. Mineral. Petrol.* **134**, 3–16 (1999).
- Krogh, T. E. Improved accuracy of U–Pb zircon ages using an air abrasion technique. *Geochim. Cosmochim. Acta* **46**, 637–649 (1982).
- Krogh, T. E. A low-contamination method for hydrothermal decomposition of zircon and extraction of U and Pb for isotope age determinations. *Geochim. Cosmochim. Acta* **37**, 485–494 (1973).
- Ballentine, C. J., Lee, D.-C. & Halliday, A. N. Hafnium isotopic studies of the Cameroon line and new HIMU paradoxes. *Chem. Geol.* **139**, 111–124 (1997).

Supplementary information is available on Nature’s World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank G. J. H. Oliver for providing the Acasta gneiss samples, and F. Corfu, D. Davis, U. Schaltegger, W. Mueller, F. Albarède and J. Patchett for comments on the manuscript. This work was supported by the NSERC, NSF and DOE.

Correspondence and requests for materials should be addressed to Y.A. (e-mail: yuria@rom.on.ca).

A complete human pelvis from the Middle Pleistocene of Spain

Juan-Luis Arsuaga*, Carlos Lorenzo*, José-Miguel Carretero†, Ana Gracia*, Ignacio Martínez*, Nuria García*, José-María Bermúdez de Castro‡ & Eudald Carbonell§

* Departamento de Paleontología, Instituto de Geología Económica, Facultad de Ciencias Geológicas, Universidad Complutense de Madrid, 28040 Madrid, Spain

† Departamento de Ciencias Históricas y Geografía, Facultad de Humanidades y Educación, Universidad de Burgos, 09071 Burgos, Spain

‡ Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, José Gutiérrez Abascal 2, 28006 Madrid, Spain

§ Laboratori d’Arqueologia, Universitat Rovira i Virgili, Plaza Imperial Tàrraco 1, 43005 Tarragona, Spain

The Middle Pleistocene site of Sima de los Huesos in Sierra de Atapuerca, Spain, has yielded around 2,500 fossils from at least 33 different hominid individuals¹. These have been dated at more than 200,000 years ago^{2–4} and have been classified as ancestors of Neanderthals^{5,6}. An almost complete human male pelvis (labelled Pelvis 1) has been found, which we associate with two fragmentary femora. Pelvis 1 is robust and very broad with a very long superior pubic ramus, marked iliac flare, and a long femoral neck. This



Figure 1 Superior view of Pelvis 1. The ischiatic spines are not reconstructed. Note the very long superior pubic ramus and the triangular shape of the pelvic inlet. Scale bar, 5 cm.

pattern is probably the primitive condition from which modern humans departed. A modern human newborn would pass through the birth canal of Pelvis 1 and this would be even larger in a female individual. We estimate the body mass of this individual at 95 kg or more. Using the cranial capacities of three specimens from Sima de los Huesos, the encephalization quotients are substantially smaller than in Neanderthals and modern humans.

A large collection of human pelvic material has been recovered from Sima de los Huesos (SH), including an almost complete pelvis (labelled Pelvis 1), two very complete coxal bones lacking the pubic bone (AT-800, AT-1004), and one less complete coxal bone that preserves the proximal half of the superior pubic ramus (Coxal I). There are also many other fragmentary specimens, including an almost complete, isolated pubic bone (AT-1006). Pelvis 1 is not distorted, although the right pubic bone is broken and isolated and the left one has a minimum point of contact with the rest of the left coxal bone and preserves only a small portion of the acetabular articular surface and floor (Figs 1 and 2). Despite this, the reconstruction of the pelvis is satisfactory, although the sagittal diameter varies (but by less than 0.5 cm) depending on the pubic orientation.

The pelves from SH are particularly important because there are few well preserved pelvic remains of non-modern human fossils. The other available non-modern (and more or less complete) pelves are A.L. 288-1 from Hadar (assigned to *Australopithecus afarensis*), Sts 14 from Sterkfontein (*Australopithecus africanus*) and the Neanderthal specimen from Kebara in Israel, dated at around 60,000 years (ref. 7).

The iliac region of Pelvis 1 exhibits well developed anterior superior iliac spines which are much deviated medially, strong iliac pillars and thick iliac tubercles (left, 31.9 mm; right, 27.1 mm), pronounced lateral iliac flare, robust and sigmoid-shaped anterior inferior iliac spines, an acetabular roof protruding as a shelf, and a well developed supra-acetabular groove. All the SH iliac remains share this pattern, which in our opinion is similar to other 'archaic' human fossils: ER 3228 (East Turkana, Late Pliocene or Early Pleistocene), OH 28 (Olduvai, Early/Middle Pleistocene), Arago 44 (France, Middle Pleistocene) and Neanderthals.

Pelvis 1 has a very long pubic bone (see Table 1). This seems to be a primitive feature for hominids, as it is also found in australopithecines, in Neanderthals and in the Middle Pleistocene specimen from Jinniushan in China⁸. However, although the pubis of Pelvis 1 is extraordinarily long, the ischium is also very long: the ischiopubic index (96.7%) is between the modern male and female means. Our tentative estimation of the ischiopubic index for the Neanderthal La Ferrassie I (108.1%) is around one standard deviation (s.d.) above the modern female mean, and Kebara yields a very high index (120%), which is 3 s.d. above the modern female mean. In AL 288-1 and Sts 14, the ischiopubic index far exceeds the modern human range⁹.



Figure 2 Ventral view of Pelvis 1. Note the marked iliac flare and the large ischiopubic frame. Scale bar, 10 cm.

The superior pubic ramus of Pelvis 1 has a very sharp pectineal crest and its superior surface is concave mediolaterally. This morphology is shared by the other five SH superior pubic rami, which, although they are all slender, never show the extremely plate-like and craniocaudally thin cross-section of most Neanderthal specimens (Fig. 3). The SH cross-sections are more similar to that of Jinniushan⁸, although the superior pubic ramus is thicker and generally shorter in modern humans, including the Skhul/Qafzeh early modern samples.

The SH Pelvis 1 is sexed as a male because of its extreme robustness, its large cotylosiatic breadth, the triangular shape of the pelvic inlet, the anterior position of the sacroiliac joint, the narrow ischial notch and its pubic morphology. Other coxal bones in the SH sample (AT-1004 and Coxal I) are less robust and show smaller cotylosiatic breadths, broader ischial notches and a more posterior position of the sacroiliac joint (which are feminine traits in modern humans⁹), whereas AT-800 is similar to Pelvis 1. The AT-1006 pubic bone exhibits a more feminine condition than Pelvis 1, having a larger non-articular pubic length (92 mm), a larger pubic body breadth, a subpubic concavity, a narrower medial aspect of the ischiopubic ramus, and even a possible ventral arc (although the bony ridge lies close to the symphyseal face, it curves away inferiorly).

The maximum breadth of the false pelvis (the bicrestal distance) in Pelvis 1 exceeds that of any previously known hominid specimen. We assign two partial femora to Pelvis 1: one large left fragment (Femur X), representing slightly more than the proximal half, and a right diaphysis (AT-432). The head of Femur X fits very well in the left acetabulum of Pelvis 1: moreover, there is an atypical remodeling of the articular surface of the acetabulum that matches a similar pattern in the femoral head. The femoral length has been estimated at between 47 and 48 cm by reconstructing the whole femur using Femur X, AT-432 and other distal femora of SH, and from the femoral head diameter by using a regression of a modern male sample from Coimbra in Portugal. From this femoral length (47–48 cm), the stature is estimated^{10,11} at between 173.3 and 179.5 cm. Using this stature range and the bicrestal breadth, the body weight was estimated¹² at between 93.1 and 95.4 kg. This weight is greater than for any other human fossil¹² and could be even greater if body mass in Neanderthals and other archaic humans were larger relative to the skeletal frame than in modern humans¹³, which is likely given the extreme thickness of the iliac pillar and the iliac tubercle. Pelvis 1 is not an atypically large individual in Sima de los Huesos. The maximum coxal height of Pelvis 1 is 240 mm, whereas the value of AT-800 is estimated at 245 mm. The acetabular height of Pelvis 1 is 60.5 mm, and that of AT-800 is 59 mm; the coxal fragments AT-835 + AT-2501 (60 mm) and AT-2350 (59 mm) have similar acetabular heights. The two SH coxal bones that were

Table 1 Measurements of Pelvis 1 and Femur X, and comparison with modern human samples

	SH	Male Z	Female Z	Males		Females	
				X	s.d.	X	s.d.
Pubic length	(102)	2.8	2.4	86.5	5.6	87.9	5.8
Non-articular pubic length (acetabulo-symphyseal length)	82	3.6	2.9	66.9	4.2	69.5	4.4
Ischial length	105.5	2.3	5.0	94.5	4.8	84.7	4.2
Ischiopubic index*	96.7	1.1	-1.3	91.7	4.7	103.8	5.4
Maximum pelvic breadth (bicrestal breadth)	340	5.9	4.5	261.7	13.3	262.4	17.3
Maximum coxal height	240	3.4	5.6	206.6	9.8	189.5	9.0
Lower iliac height	(57)	0.0	-1.2	56.8	5.5	61.1	6.3
Vertical acetabular diameter	60	1.7	3.7	55.2	2.8	49.9	2.7
Transverse inlet diameter	138	2.5	0.9	123.1	6.0	130.4	8.4
Minimum biacetabular breadth	133.5	3.8	2.4	109.2	6.4	114.5	7.9
Transverse midplane diameter (bispinal breadth)†	(108)	2.7	0.5	89.0	7.0	103.8	9.4
Transverse outlet diameter (bituberous breadth)	135	2.3	1.0	110.3	10.7	122.9	12.1
Maximum transverse diameter of the midplane	121	3.0	1.0	99.5	7.2	115.5	9.1
Sagittal inlet diameter	121	1.6	0.9	105.6	9.6	112.6	10.0
Obstetrical conjugate	109.4	1.0	0.2	99.5	9.9	107.7	9.6
Sagittal midplane diameter‡	133	2.0		117	8.0		
Sagittal outlet diameter	126.5	2.1	1.2	109.2	8.1	116.1	8.9
Circumference inlet‡	405	1.5		370	24		
Circumference midplane‡	347	3.4		297	15		
Circumference outlet‡	395	3.8		324	19		
Subpubic angle	76	1.1	-0.4	66.8	8.7	79.4	8.3
Divergence angle of the iliac blades§	134	3.8	3.4	98.7	9.3	99.2	10.3
Sacral breadth‡	111.1	0.7		106	7.0		
Proximal width of the femur	111.9	4.3	6.1	90.8	4.9	82.6	4.8
Relative proximal width of the femur	23.3	1.9	2.4	20.7	1.4	20.4	1.2

Measurements are given in millimetres and degrees. Parentheses indicate estimated values. Definition of variables and data are from refs 9, 22. All measurements are from skeletons of known sex in the collection of the Instituto de Antropologia da Universidade de Coimbra, Portugal.

* Pubic length/ischial length $\times 100$.

† With reconstructed ischiatic spines as large and pointing, although they could have been blunt and shorter (but never longer).

‡ Definition of the variables and comparative data in ref. 23 for a male pooled sample of six modern populations.

§ The divergence angle is taken in the outer margins of the iliac tubercles, instead of in the inner margins. The thickness of the iliac tubercles in Pelvis 1 exaggerates the value of the angle.

|| Proximal width/maximum length. We have used an estimated maximum length of 480 mm for Femur X.

X = mean; Male Z = (SH value - male mean)/male s.d.; Female Z = (SH value - female mean)/female s.d.

sexed as females yielded smaller acetabular heights (Coxal I, 50 mm; AT-1004, 54.5 mm).

Cranial capacity has been directly measured in two SH specimens: Cranium 5 (1,125 cm³) and Cranium 4 (1,390 cm³); in a third individual, Cranium 6, it can be accurately estimated at 1,220 cm³ (ref. 14, 15). Because Cranium 5 has one of the smallest cranial capacities in Middle Pleistocene sample except *Homo erectus* and that of Cranium 4 is the largest for the period, these figures can be considered close to the ends of the SH range (which probably has a mean smaller than in modern humans). Using a cranial capacity value of 1,390 cm³, and a body-weight range of 93.1–95.4 kg, the encephalization quotient would be 3.7–3.8, well below the figure for modern humans (5.3–5.4) and also below the Neanderthal value (5.0 $\pm$ 0.2)¹². Using the cranial capacity of Cranium 5, the encephalization quotient would be even lower (3.0–3.1). The encephalization quotient is not meaningful when closely related species with widely differing body mass are compared. In these cases, it is more informative to analyse the encephalization trends. Because Neanderthals maintained the same body mass as their European Middle Pleistocene ancestors (or even became slightly smaller) while their

brains expanded, we conclude that Neanderthal encephalization was greater than in their ancestors.

The transverse diameter of the inlet is large in Pelvis 1 but the bicrestal breadth is much greater, so the iliac blades markedly flare laterally. The biacetabular breadth and the transverse diameter of the outlet (bituberous diameter) are considerable. Both ischiatic spines are broken and have been reconstructed, yielding a wide transverse diameter of the midplane (bispinal breadth). The maximum transverse diameter of the midplane (taken at the level of the ischiatic spines but on the internal walls of the pelvic cavity) was also large. Indeed, Pelvis 1 is very wide in all transversal diameters: at the iliac crest, inlet, midplane and outlet levels.

A broad pelvis with a large biacetabular breadth is biomechanically disadvantageous for the lateral support system of the hip, because it creates a long abductor load arm. It is compensated for in Pelvis 1 by a long femoral neck (the abductor power arm) and marked iliac flaring that verticalizes the line of action of the abductor muscles gluteus medius, gluteus minimus and tensor fasciae latae, which arise from the iliac blade and insert in the greater trochanter. Because such a large biacetabular breadth is obstetrically wasteful (as we will see later), there must be another (non-obstetrical and non-biomechanical) explanation. One possibility is that the extreme pelvic breadth of Pelvis 1 would represent an adaptation to a cold climate, minimizing the surface-to-volume ratio of the body cylinder¹⁴. However, our interpretation is that the pattern of a broad pelvis, a long femoral neck and marked iliac flaring is a shared primitive character already present in *Australopithecus afarensis* (as seen in the A.L. 288-1 specimen), in which case it should be also found in the early *Homo* fossils. However, based on the juvenile and very fragmentary WT 15000 pelvis (West Turkana, Early Pleistocene), the 'adult' bicrestal breadth of *Homo ergaster* has been estimated as narrow¹⁶. In our opinion, the East African ER 3228 and OH 28 coxal bones are so similar in all the preserved regions to the SH coxal bones, also showing a marked lateral iliac flare, that the inferred narrow bicrestal breadth for WT 15000 might be an error. A narrow pelvis could be a unique modern human condition.

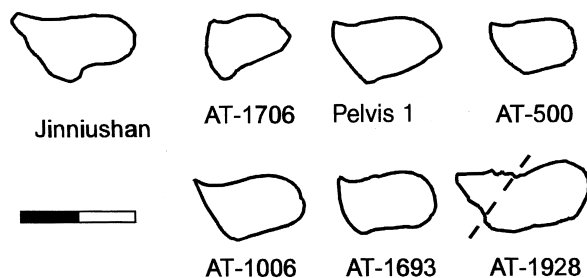


Figure 3 Cross-sections of the superior pubic rami at the highest point of the obturator foramen in Jinniushan⁸ and six SH specimens. Anterior is to the right, posterior to the left. In AT-1928, the portion to the left of the dashed line is slightly displaced. AT-1693 is an immature pelvis with an acetabulo-symphyseal length of 84 mm. Scale bar, 2 cm.

The sagittal diameters of the pelvic inlet, outlet and midplane of the Pelvis 1 birth canal are also large. However, the pubic symphysis of this specimen is very thick (24.8 mm) and dorsally convex (looking like 'pubic promontorium'), so the obstetrical conjugate is only 1 s.d. above the modern male mean (but it is still larger than 84% of modern males, and is close to the modern female mean). The long superior pubic ramus of Pelvis 1 places the pubic symphysis well in front of the hip joints, a morphology previously observed in the Kebara pelvis¹⁷. However, the acetabula are no closer to the sacroiliac joints than in modern human males, as the lower iliac height shows.

Although Pelvis 1 is from a male individual, its birth canal could be easily negotiated by a fetus of modern human dimensions at term (biparietal diameter, 9.3 ± 1.1 mm; suboccipitobregmatic diameter, 9.4 ± 1.2 mm)¹⁸, even if the pelvic diameters were diminished by 5 mm by the soft tissues covering the inner pelvis. In modern males and females, the smallest diameter of the pelvic canal is usually the bispinal diameter, which defines the obstetric plane of least dimensions and constrains the descent of the fetus through the pelvis during birth¹⁹. Because the bispinal diameter is greater in females than in males in all human populations and in pongids²⁰, it would also be greater in the SH females, so a fetal head at least similar to, and probably bigger than, that of modern human fetuses would pass through an SH female pelvic cavity. This would mean that the percentage of adult brain size at birth would be higher than in modern humans, as the SH adult brain size was probably slightly inferior.

Finally, in Pelvis 1, the greatest dimension of the pelvic inlet is the transverse diameter, but in the midplane it is the sagittal diameter, so the birth canal is twisted. In modern humans, both sexes show twisted birth canals and this was probably the case in the SH population. Because the greatest fetal diameters adapt to the greatest pelvic dimensions, parturition would be as in modern humans, with the fetus rotating within the birth canal.

One enigmatic and controversial aspect of the palaeodemography of the Middle Pleistocene and Neanderthal samples is the apparent absence of aged individuals²¹. However, according to modern human patterns of remodelling of the symphyseal and auricular surfaces, Pelvis 1 would be older than 35 years, and AT-2500 (an isolated pubic body) would be more than 45 years old. The patterns of symphyseal and avicular remodelling could have been different in Middle Pleistocene, but this could also have been the case of the most commonly used ageing method, dental wear.

Received 11 November 1998; accepted 19 March 1999.

1. Arsuaga, J. L. *et al.* Sima de los Huesos (Sierra de Atapuerca, Spain). The site. *J. Hum. Evol.* **33**, 109–127 (1997).
2. García, N., Arsuaga, J. L. & Torres, T. The carnivore remains from the Sima de los Huesos Middle Pleistocene site (Sierra de Atapuerca, Spain). *J. Hum. Evol.* **33**, 155–174 (1997).
3. Bischoff, J. L. *et al.* Geology and preliminary dating of the hominid-bearing sedimentary fill of the Sima de los Huesos Chamber, Cueva Mayor of the Sierra de Atapuerca, Burgos, Spain. *J. Hum. Evol.* **33**, 129–154 (1997).
4. Cuenca-Bescós, G., Laplana Conesa, C., Canudo, J. I. & Arsuaga, J. L. Small mammals from Sima de los Huesos. *J. Hum. Evol.* **33**, 175–190 (1997).
5. Arsuaga, J. L., Martínez, I., Gracia, A., Carretero, J. M. & Carbonell, E. Three new human skulls from the Sima de los Huesos Middle Pleistocene site in Sierra de Atapuerca, Spain. *Nature* **362**, 534–537 (1993).
6. Arsuaga, J. L., Martínez, I., Gracia, A. & Lorenzo, C. The Sima de los Huesos crania (Sierra de Atapuerca, Spain). A comparative study. *J. Hum. Evol.* **33**, 219–281 (1997).
7. Rak, Y. In *Le squelette Moustérien de Kébara 2* (eds Bar Yosef, O. & Vandermeersch, B.) 147–156 (CNRS, Paris, 1991).
8. Rosenberg, K. In *Neandertals and Modern Humans in Western Asia* (eds Akazawa, T., Aoki, K. & Bar-Yosef, O.) 367–379 (Plenum, New York, 1998).
9. Arsuaga, J. L. & Carretero, J. M. Multivariate analysis of the sexual dimorphism of the hip bone in a modern population and in early hominids. *Am. J. Phys. Anthropol.* **93**, 241–257 (1994).
10. Feldsman, M. R., Kleckner, J. G. & Lundy, J. K. Femur/stature ratio and estimates of stature in mid- and late-Pleistocene fossil hominids. *Am. J. Phys. Anthropol.* **83**, 359–372 (1990).
11. Trotter, M. & Gleser, G. C. Estimation of stature from long bones of American Whites and Negroes. *Am. J. Phys. Anthropol.* **10**, 463–514 (1952).
12. Ruff, C. B., Trinkaus, E. & Holliday, T. W. Body mass and encephalization in Pleistocene *Homo*. *Nature* **387**, 173–176 (1997).
13. Kappelman, J. They might be giants. *Nature* **387**, 126–127 (1997).
14. Arsuaga, J. L. *et al.* Size variation in Middle Pleistocene humans. *Science* **277**, 1086–1088 (1997).
15. Lorenzo, C., Carretero, J. M., Arsuaga, J. L., Gracia, A. & Martínez, I. Intrapopulation body size variation and cranial capacity variation in Middle Pleistocene humans. The Sima de los Huesos sample (Sierra de Atapuerca, Spain). *Am. J. Phys. Anthropol.* **106**, 19–33 (1998).

16. Ruff, C. B. & Walker, A. In *The Nariokotome Homo erectus Skeleton* (eds Walker, A. & Leakey, R. E. F.) 234–265 (Springer, Berlin, 1993).
17. Rak, Y. In *Species, Species Concepts, and Primate Evolution* (eds Kimbel, W. H. & Martin, L. B.) 523–536 (Plenum, New York, 1993).
18. Abitbol, M. M. *Birth and Human Evolution* (Bergin & Garvey, London, 1996).
19. Walrath, D. E. & Glantz, M. M. Sexual dimorphism in the pelvic midplane and its relationship to Neanderthal reproductive patterns. *Am. J. Phys. Anthropol.* **100**, 89–100 (1996).
20. Tague, R. G. Commonalities in dimorphism and variability in the anthropoid pelvis, with implications for the fossil record. *J. Hum. Evol.* **21**, 153–176 (1991).
21. Trinkaus, E. Neanderthal mortality patterns. *J. Archaeol. Sci.* **22**, 121–142 (1995).
22. Serra, J. A. A pelve nos portugueses. Morfologia da pelve no homem. *Contrib. Estudo Anthropol. Port.* **15**, 1–173 (1998).
23. Tague, R. G. Sexual dimorphism in the human bony pelvis, with a consideration of the Neanderthal pelvis from Kebara Cave, Israel. *Am. J. Phys. Anthropol.* **88**, 1–21 (1992).

Acknowledgements. We thank G. Manzi, Y. Rak and I. Hershkovitz for comments on the manuscript, and J. Trueba for providing the photographs. C.L. received a grant from Ayuntamiento de Madrid in the Residencia de Estudiantes. This research is partly funded by the Junta de Castilla y León, the Dirección General de Enseñanza Superior de Spain (PB93-0066-C03), Comunidad de Madrid (CAM 06/0037/1997) and Unidades Asociadas (CSIC).

Correspondence and requests for materials should be addressed to J.-L.A. (e-mail: azara@eucmax.sim.ucm.es).

Relative risk of extinction of passerine birds on continents and islands

Lisa L. Manne*, Thomas M. Brooks*† & Stuart L. Pimm*

* Department of Ecology and Evolutionary Biology, 569 Dabney Hall, The University of Tennessee, Knoxville, Tennessee 37996, USA

† Department of Biological Sciences and Center for Advanced Spatial Technologies, 12 Ozark Hall, University of Arkansas, Fayetteville, Arkansas 72701, USA

Greater numbers and higher proportions of recent species extinctions have been on islands rather than on continents. In contrast, predictions of massive future extinctions stem from the current clearing of continental, tropical forests¹. For instance, since 1600, 97 out of 108 bird extinctions have been on islands². However, 452 of the total 1,111 species currently considered to be threatened are continental³. Island flora and fauna are uniquely vulnerable to the human introduction of previously absent predators, diseases and other menaces⁴, whereas species on continents are not so ecologically naive. So could predictions of future continental extinctions based on island histories be exaggerated¹? Most threatened species have small geographic ranges⁵, and the ranges of island species are inevitably smaller than those of continental species. For a given range size, how do the proportions of threatened island and continental species compare? Here we compile the ranges of the passerine (perching) birds of the Americas. Corrected for range size, continental species are more—not less—likely to be threatened. We use this unexpected vulnerability of continental species with small ranges to produce a map showing where species losses might occur in the long term.

To separate the effects of range size from those of island compared to continental distribution, we calculate the breeding ranges (henceforth, just 'ranges') of all the passerine birds of the Americas and their associated islands. This complete enumeration of New World passerines constitutes nearly a quarter of the ~10,000 bird species of the world. Some of the continental species inhabit montane habitat 'islands' isolated by a 'sea' of lowland habitats. To investigate whether these montane species suffer specific levels of threat, we separate them from lowland species.

The median range sizes for lowland, montane and island species are 1,000,000 km², 210,000 km² and 7,600 km², respectively (Table 1). Threatened species are those considered likely to have a "high probability of extinction in the medium-term future"³. Their corresponding median range sizes are 72,000 km², 53,000 km² and 450 km². A larger fraction of island species than of continental

Table 1 All (and threatened) species by range size and geographic location

Habitat	Range size in square kilometres						
	<100	100–1,000	1,001–10,000	10,001–100,000	100,001–1,000,000	1,000,001–10,000,000	10,000,001–100,000,000
Lowland	8 (4)	7 (2)	25 (16)	108 (37)	665 (42)	787 (5)	68 (0)
Montane	0 (0)	8 (1)	42 (16)	123 (23)	286 (5)	37 (0)	4 (0)
Island	6 (3)	21 (5)	29 (3)	46 (6)	17 (1)	0	0

species is at risk of extinction. Of the 2,168 continental species, 106 out of 1,668 lowland species (6%), 45 out of 500 montane species (9%) and 18 out of 119 island species (15%) are threatened.

Figure 1 shows threatened species as a proportion of total species. The data presented include all of the New World passerines and not, for instance, a random sample of all passerines. Thus, we cannot draw statistical inferences about the levels of threat for all passerines. Our results are complete for the New World passerines, not a sample of them. As such, they are subject only to the uncertainties of our measurements of range size and the ways in which the species are classified.

Figure 1 excludes the 15 lowland, 8 montane and 27 island species with ranges smaller than 1,000 km². At these range sizes, the proportions of threatened species are high, but uncertain, on both islands and continents. For these small numbers, whether or not one includes with 'threatened' species those deemed to be threatened on the basis of range size alone, or those typically rare species for which there are insufficient data, alters the proportions considerably. Despite these uncertainties, there is no evidence that island species are more vulnerable than those on continents. Perhaps this should be expected: tiny ranges should make species vulnerable to habitat loss, hunting and other threats, wherever the species live.

We found an unexpected result for range sizes between 1,000 and 100,000 km². A much higher proportion of lowland than of either montane or island species is threatened. (For this subset of the data, the inclusion of data-deficient species yields no discernible differences in Fig. 1.)

At the largest ranges, greater than 100,000 km², we expect that the proportions of threatened species should be much lower and similar for all groups of species. Large islands often house the same sets of hazards as continents, threats that are missing from smaller islands. The proportions are indeed similar. No island species have ranges larger than 250,000 km², so no comparisons are possible for this range size.

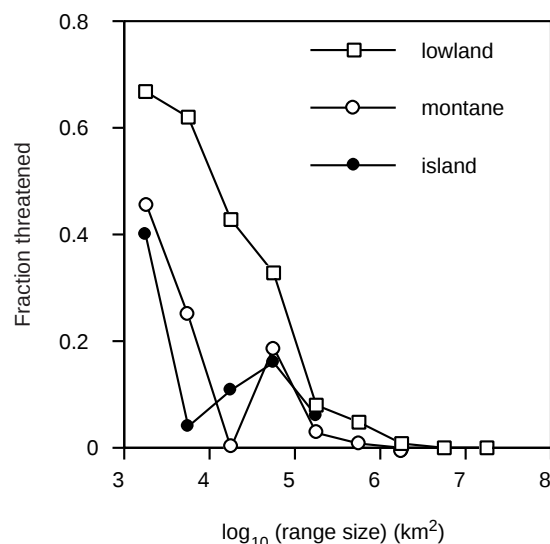


Figure 1 The fraction of threatened lowland, montane and island passerines in each range-size division. Open squares, lowland; open circles, montane; filled circles, island.

The unexpected result for range sizes below 100,000 km² might have several explanations; here we present two. First, the 10,000 years of human occupation of America's islands may have already driven a large percentage of the most sensitive species to extinction. The Caribbean islands have certainly lost species. Five species of passerine birds and a further 35 non-passerines are known only from sub-fossil remains⁶. Elsewhere, even well-worked archaeological sites find only half the extinct species, so these numbers must be assumed to underestimate their true total⁷. If a large number of already-extinct species had survived to the present day, their inclusion as threatened species in Fig. 1 would bring the island results nearer to those for the continental lowlands.

Species might also have been lost historically from montane regions. Most of the montane species are along the continent's western rim from the Sierra Madre Occidental of Mexico, through central America and southwards along the Andes. The human population of this region is predominantly montane, as exemplified by the elevations of eight of its thirteen cities with more than 1 million people: Guadalajara (1,579 m), Mexico City (2,240 m), Puebla (1,525 m), Guatemala City (1,478 m), Managua (59 m), Barranquilla (0 m), Bogota (2,640 m), Cali (610 m), Medellin (1,538 m), Quito (2,841 m), Guayaquil (150 m), Lima (152 m), La Paz (3,901 m)⁸. This implies that the montane species inhabit the same places that the people preferentially occupy, leaving them vulnerable to human-caused extinction.

Our second explanation relates to the relative population densities of the species. Consider Jamaica, an island of 11,000 km² that houses 15 passerine endemics. None is threatened, and nine are common across most of the island's habitats⁹. Similar observations pertain to the 18 passerines endemic to the Galapagos; two are threatened, but most are common¹⁰. Montane species with small ranges are also often common in those ranges¹¹. These examples of common species with small ranges have no equivalents in continental lowlands. There, species with such small ranges are almost always very rare within those ranges. A reasonable explanation for this phenomenon is competitive release on islands¹² (and, by extinction, habitat islands). With fewer competitors, island species are able to attain higher densities and concomitant broader ecological niches¹³.

Across many taxa and many areas, widespread species are locally more dense than range-restricted species¹⁴. If our second explanation holds, we predict that the same relationship will still be true of island and montane species, but their densities would be greater than those of lowland continental species at a given range size. Thus, the increased threat for lowland continental species might be because those with small ranges also have low densities.

Figure 1 shows that lowland continental species with ranges < 100,000 km² are disproportionately threatened, as are most species with ranges < 10,000 km². Figure 2a combines these two classes as a hemisphere-wide prediction of where threatened species should occur. These species are concentrated in the Lesser Antilles, the Galapagos, Costa Rica and Panama, the Andes and the Atlantic coast forests of Brazil.

Figure 2b maps the actual distribution of threatened passerines. (This map is similar to that derived independently for all threatened passerines and non-passerines¹⁵.) Threatened species are most numerous along the Atlantic seaboard of South America. Extensive deforestation of the coastal forests¹⁶ and, to the south, clearing of the pampas for agriculture¹⁷, has endangered many of the region's

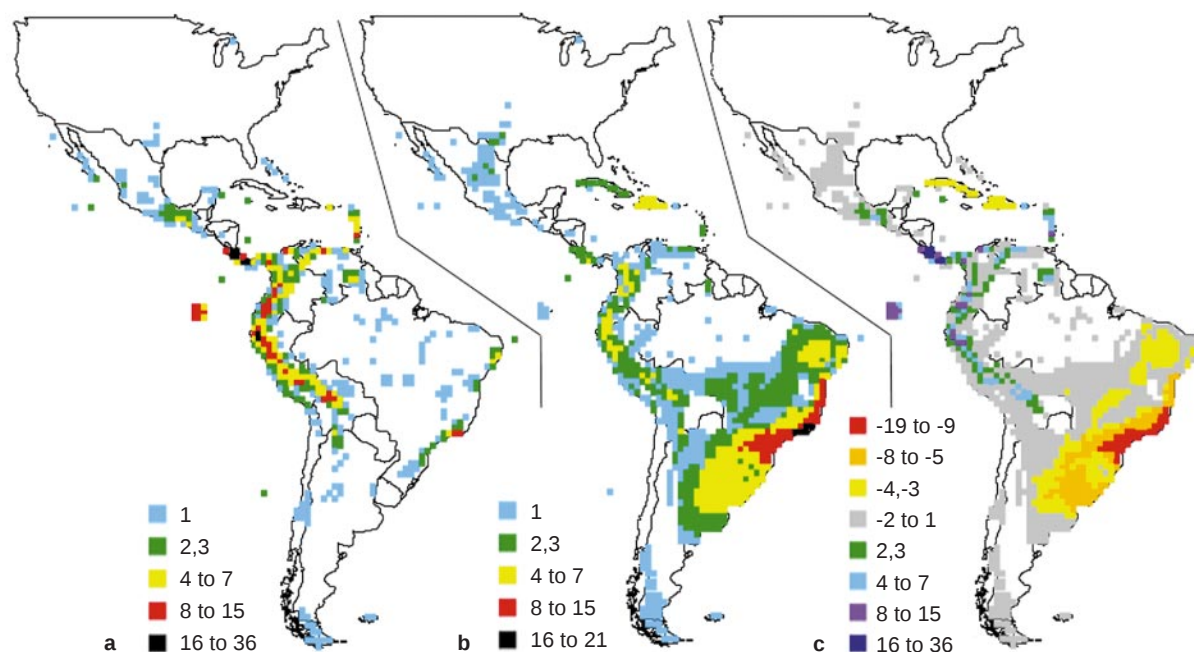


Figure 2 Distributions of passerines with small ranges and of threatened passerines, and differences between them. **a**, The numbers of lowland passerines with ranges <100,000 km² plus all species with ranges <10,000 km². **b**, The numbers of threatened passerines. **c**, The difference in numbers between **a** and **b**. White areas are locations holding neither observed nor expected threatened species. Areas that are white in **a** but coloured in **b**, or vice versa, will

be coloured in **c**. Grey areas indicate a near-match between our model and the actual locations of threatened species. Red, orange or yellow areas are locations where there are more threatened species than our model predicts. Green, blue or purple areas are locations where there are fewer threatened species than our model predicts.

species. There are lower concentrations of threatened species along the Andes and on Caribbean islands.

If there were no differences between observed and predicted numbers of threatened species, it would suggest that human impacts apply uniformly across the continents. In Fig. 2c, we subtract the actual numbers of threatened species from the predicted numbers. White areas are where there are neither predicted nor observed threatened species. The grey areas, where our model nearly matches the number of threatened species, are extensive. Nonetheless, there are more threatened species than our model predicts along the southern Atlantic coast of South America (shown in yellow to red). The differences arise from the collision of areas with many small-ranged species and dense and expanding human populations. Areas with fewer threatened species than predicted are concentrated in the Andes, Central America, the Lesser Antilles and southern Mexico (shown in green to purple).

Areas with more threatened species than expected indicate where extinctions will occur first. These provide the most immediate conservation priorities where, by force of circumstance, management is reactive and intensive. Might areas with fewer than expected threatened species be where conservation efforts should be directed in the long-term? Certainly, these areas need not hold many threatened species in the future, because the spread of human impacts is impossible to guess. However, other things being equal, areas rich in not-yet-threatened species with small ranges house the most vulnerable targets for conversion to threatened status. The unexpected vulnerability of range-restricted, lowland continental species motivates proactive management for the areas where they occur. Here, ecosystem protection now would obviate the need for intensive management later. □

Methods

'Island' species breed only on islands, 'continental' species are all other species. Of the latter, 'montane' species do not occur below 1,000 m and the remainder are 'lowland' species. For widespread continental species, we digitized the range

maps from the regional sources^{10,18–25} onto a one-degree latitude-longitude grid. We then summed the areas of these cells for every species. This method has potentially large errors for species with the smallest ranges (<4 cells, roughly 40,000 km² in area). For them, we employed fine-resolution topographic and vegetation maps. Using the particular locality and elevational range requirements of each species, we estimated the maximum possible area of the habitat type. For island species, for which maps did not delineate smaller regions, we used the total island area as the range of the species. We placed six species known from one specimen or one locality into our smallest range class (<100 km²). All of these approaches will overestimate the true range to varying degrees. Not all of the area will be suitable habitat. Such errors probably apply equally across the divisions of our data and do not obviously lead to consistent biases.

Collar *et al.*³ classify 184 species as 'threatened'. Of these species, 15 are 'D2s'—those threatened solely on the basis of small range size alone. (Nine are continental and six are insular; all but three have ranges less than 100 km²; these three are from scattered locations.) Two more are classed as 'DD'—data-deficient species about which not enough is known to pass judgement.

Might small-ranged continental species inevitably be threatened, because it is the threats that have so reduced the species' ranges? Generally, the range maps do not reflect range contractions due to recent massive destruction of tropical forests. Thus, species with small ranges are probably those with historically small ranges, although very little is known about range changes.

We checked whether taxonomic decisions (splitting or lumping) could be applied inconsistently to island and continental species and so be responsible for the unexpected differences in relative threat. We re-analysed the data using Sibley and Monroe, the only globally standard taxonomy^{26,27}. They lump some species, but the impact of these changes on Fig. 1 is not visually discernible.

For species names, range sizes, categories and taxonomic decisions, see Supplementary information.

Received 11 September 1998; accepted 5 March 1999.

1. Pimm, S. L. *et al.* The future of biodiversity. *Science* **269**, 347–350 (1995).
2. Johnson, T. H. & Stattersfield, A. J. A global review of island endemic birds. *Ibis* **132**, 167–180 (1990).
3. Collar, N. J., Crosby, M. J. & Stattersfield, A. J. *Birds to Watch 2* (Smithsonian Institution Press, Washington, D.C., 1994).
4. Pimm, S. L. *The Balance of Nature?* (University of Chicago Press, Chicago, 1991).
5. Stattersfield, A. J. *et al.* *Endemic Bird Areas of the World* (Smithsonian Institution Press, Washington,

- DC, 1998).
- Milberg, P. & Tyrberg, T. Naive birds and noble savages—a review of man-caused prehistoric extinctions of island birds. *Ecography* **16**, 229–250 (1993).
 - Pimm, S. L., Moulton, M. P. & Justice, L. J. in *Extinction Rates* (eds Lawton, J. H. & May, R.) 75–87 (Oxford University Press, New York, 1995).
 - Goode's World Atlas 18th edn (ed. Espenshede, E. Jr) (Rand McNally, Chicago, 1990).
 - Lack, D. *Island Biology Illustrated by the Land Birds of Jamaica* (Blackwell, London, 1976).
 - Castro, I. & Phillips, A. *A Guide to the Birds of the Galapagos Islands* (Princeton University Press, Princeton, 1996).
 - Stotz, D. F. et al. *Neotropical Birds, Ecology and Conservation* (Univ. of Chicago Press, Chicago, 1996).
 - MacArthur, R. H., Diamond, J. M. & Karr, J. Density compensation in island faunas. *Ecology* **53**, 330–342 (1972).
 - Green, R. E. & Hiron, G. J. M. in *Bird Population Studies* (eds Perrins, C., Lebreton, J.-D. & Hiron, G.) 683 (Oxford University Press, Oxford, 1991).
 - Gaston, K. J., Blackburn, T. M. & Lawton, J. H. Interspecific abundance-range size relationships: an appraisal of mechanisms. *J. Anim. Ecol.* **66**, 579–601 (1997).
 - Collar, N. J., Wedge, D. C. & Long, A. J. in *Ornithological Monographs*, 48, *Studies in Neotropical Ornithology Honoring Ted Parker* (ed. Remsen, J. V. J.) 237–260 (American Ornithologists Union, Winfield, KS, 1997).
 - Brooks, T. M. & Balmford, A. Atlantic forest extinctions. *Nature* **380**, 115 (1996).
 - Bucher, E. & Nore, M. in *ICBP Technical Publications*, 7, *Ecology and Conservation of Grassland Birds* (ed. Goulet, P.) 71–79 (Cambridge University Press, Cambridge, 1998).
 - Ridgely, R. S. & Tudor, G. *The Birds of South America Vol. 1* (Univ. of Texas, Austin, TX, 1989).
 - Ridgely, R. S. & Tudor, G. *The Birds of South America Vol. 2* (Univ. of Texas, Austin, TX, 1994).
 - Raffaele, H. et al. *A Guide to the Birds of the West Indies* (Princeton University Press, Princeton, NJ, 1998).
 - Howell, S. N. G. & Webb, S. A. *A Guide to the Birds of Mexico and Northern Central America* (Oxford University Press, New York, 1995).
 - Stiles, F. G., Skutch, A. F. & Gardner, D. *A Guide to the Birds of Costa Rica* (Cornell University Press, Ithaca, 1989).
 - Godfrey, W. E. *The Birds of Canada* (National Museum of Canada, Ottawa, 1986).
 - Price, J., Droege, S. & Price, A. *The Summer Atlas of North American Birds* (Academic, New York, 1995).
 - Ridgely, R. S. & Gwynne, J. A. *A Guide to the Birds of Panama* (Princeton University Press, Princeton, 1989).
 - Sibley, C. G. & Monroe, B. L. J. *Distribution and Taxonomy of Birds of the World* (Yale University Press, New Haven, 1990).
 - Sibley, C. G. & Monroe, B. L. J. *A Supplement to Distribution and Taxonomy of Birds of the World* (Yale University Press, New Haven, 1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank C. G. Anderson, K. Balent, N. Collar, J. Gittleman, R. Green, A. Mayer, M. P. Nott, B. Powell, G. Russell, W. Simms, J. Smith, A. Stattersfield and C. Wilder. This work was supported by a fellowship to S.L.P. from the Pew Foundation for Conservation and the Environment and by the Center for Conservation, Stanford University.

Correspondence and requests for materials should be addressed to S.L.P. (e-mail: StuartPimm@aol.com).

Non-commutativity in the brain

Douglas B. Tweed*, Thomas P. Haslwanter†, Vera Happe‡ & Michael Fetter‡

* Departments of Physiology and Medicine, 1 King's College Circle, University of Toronto, Toronto M5S 1A8, Canada

† Department of Neurology, University Hospital, Frauenklinikstrasse 26, Zurich 8091, and Department of Physics, ETH Hoenggerberg, Zurich 8093, Switzerland

‡ Department of Neurology, University of Tübingen, Hoppe-Seyler-Strasse 3, Tübingen 72076, Germany

In non-commutative algebra, order makes a difference to multiplication, so that $a \times b \neq b \times a$ (refs 1, 2). This feature is necessary for computing rotary motion, because order makes a difference to the combined effect of two rotations^{3–6}. It has therefore been proposed that there are non-commutative operators in the brain circuits that deal with rotations, including motor circuits that steer the eyes, head and limbs^{4,5,7–15}, and sensory circuits that handle spatial information^{12,15}. This idea is controversial^{12,13,16–21}: studies of eye and head control have revealed behaviours that are consistent with non-commutativity in the brain^{7–9,12–15}, but none that clearly rules out all commutative models^{17–20}. Here we demonstrate non-commutative computation in the vestibulo-ocular reflex. We show that subjects rotated in darkness can hold their gaze points stable in space, correctly computing different final eye-position commands when put through the same two rotations in different orders, in a way that is unattainable by any commutative system.

The vestibulo-ocular reflex (VOR) is perhaps the simplest of the many neural responses to rotary motion: sensors in the inner ear

measure head velocity and send commands to the eye muscles, moving the eyes in the opposite direction when the head turns, so as to keep the eyeballs from rotating relative to space⁶. By stabilizing our retinal images, the VOR allows us to see while moving.

An ideal VOR must be capable of non-commutative operations. This point has been argued before^{5,13}, but it is perhaps more clearly illustrated by the thought experiment in Fig. 1, which shows that the VOR, if it is to keep the gaze on target, must compute different final eye-position commands when a person undergoes the same two rotations in different orders. Starting by looking at a space-fixed target 30° to the left, a subject who turns first 10° counterclockwise (CCW) and then 60° left must end up looking 30° right and 5° up to stay on target (Fig. 1a). A subject who turns first 60° left and then 10° CCW must end up looking to the right and down (Fig. 1b).

Surprisingly, this ideal behaviour is not predicted by most models of the neural circuitry underlying the VOR. For example, Fig. 2 shows a computer simulation of an influential theory¹⁹. This model elegantly explains many features of the VOR, but as the simulation shows, it fails to keep the eye on target when put through the rotations from Fig. 1. Regardless of the order of rotations, it leaves the eye in the same final position relative to the head. Relative to

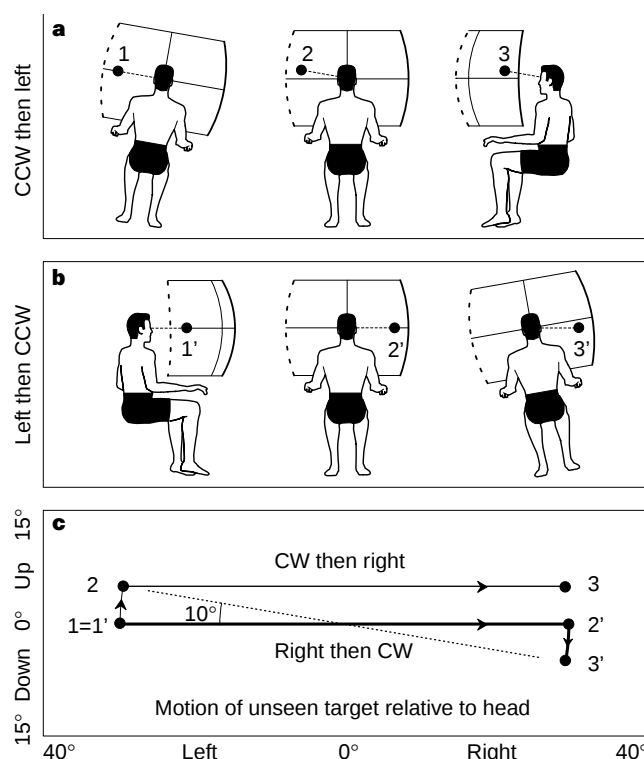


Figure 1 Why an ideal VOR must be non-commutative. **a**, Imagine yourself in a rotary capsule, tilted 10°, right ear down, as in position 1. Through a window you view a space-fixed target (black disc) 30° to the left relative to your head. The lights go out, but you are to maintain fixation on the unseen target as you rotate, first 10° counterclockwise (CCW) into position 2, then 60° left into position 3. **b**, The same rotations in the opposite sequence: starting upright, again looking 30° left relative to your head, you turn 60° left then 10° CCW. **c**, In both sequences, the space-fixed target starts at location 1 = 1' relative to your head. If you first rotate 10° CCW, the world turns 10° CW relative to you, so the target, 30° left, swings up 10° × sin(30°) = 5°, to location 2. Your subsequent rotation 60° left moves the target 60° right to location 3. But if you rotate first 60° left and then 10° CCW, the target moves via 2' to 3'. Owing to the non-commutativity of rotations, the final locations of the target, 3 and 3', differ by 2 × 10° × sin(30°) = 10°. Hence your VOR, if it is to keep your eyes on the target, must be non-commutative: it must compute different final eye-position commands, depending on the sequence of body rotations. In all figures, heavy lines depict motion when the subject turns first left then CCW; light lines show the opposite order.

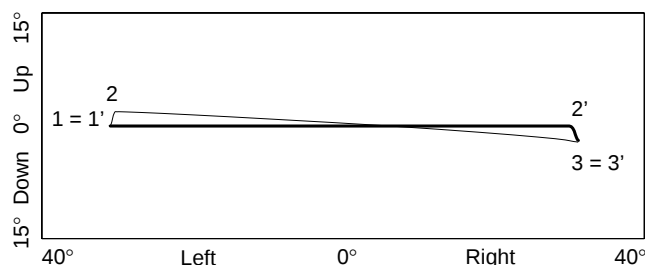


Figure 2 Computer simulation of a VOR model¹⁹ in which all neural processing is commutative. The plot shows the eye's predicted motion relative to the head when a simulated subject is put through the two sequences of rotations from Fig. 1. By the time the motions are over, the heavy and light lines coincide: the final eye position does not depend on the sequence of body rotations, and hence the eye does not stay on target.

space, then, the eye is incorrectly positioned, pointing above or below the target.

Figure 3a shows that, contrary to most theories, the actual VOR closely matches the ideal, non-commutative behaviour depicted in Fig. 1c, yielding different final eye positions depending on the sequence of rotations. In this subject, the difference (δ) between the average final eye positions, 3 and 3', was 9.0° vertical, as compared to the theoretical ideal of 10° shown in Fig. 1c. (This ideal of course depends on the rotations used: had they been larger, the ideal δ would have been larger.) Across all five of our subjects, δ averaged 10.3° (range 7.4–12.6°), and was significant (*t*-test, $P < 0.001$) for each individual. The time plots in Fig. 3b confirm that the distinct final eye positions were maintained: 3 s after the head rotations were over, δ was still 8.5° in this subject, and the average over all participants was 9.4° (range 6.5–13.9°). When the leftward and CCW head rotations in these experiments were replaced by rightward and clockwise (CW), the results were similar, with δ averaging 10.1° (range 7.6–14.3°), significant ($P < 0.001$) in each subject.

Why does the model in Fig. 2 fail to produce the necessary non-commutativity? Is it not a trivial matter to spin the eye at the same speed as the head and in the opposite direction, thereby holding it on target? Not quite. The eye-velocity command in the model is sound: it is computed in the brainstem, by taking head-velocity signals from the inner ear and multiplying them by approximately -1 , so it opposes the head motion as required. The problem is in the neural signals that code eye position. To hold the eye stable after the head rotation is finished, the model generates an eye-position command by integrating (in the calculus sense) the head-velocity signal multiplied by -1 . Unfortunately, an integrator stores no memory of the order of its inputs, so it is bound to behave commutatively: whether the head rotates first left and then CCW or first CCW and then left, it is all the same to an integrator, and so the final eye position in the model is also the same. Any model that uses solely commutative operators to transform vestibular input into eye-position commands will make the same error.

It has been proposed that non-commutative effects in ocular control need not imply non-commutative computations in the brain, because some of these effects can be explained by the geometry of the eye muscles, in particular the fact that the muscles run through pulleys in the orbital wall^{22,23}. Pulleys do have major implications for eye control^{12,19–23}, but neither they nor any other muscle geometry can explain the non-commutativity in Fig. 3. In fact, the model in Fig. 2 incorporates pulleys, but it remains commutative. Whatever the muscle geometry, there is always a correspondence between final, steady-state eye position and motor-neuron firing: to hold different eye positions, you need different firing rates^{6,24}. Even if the pulleys themselves are adjusted by a separate set of motor neurons, final eye position will still be determined by the total motor-neuron pool. The distinct final eye positions in Fig. 3 therefore imply that the final activities of the

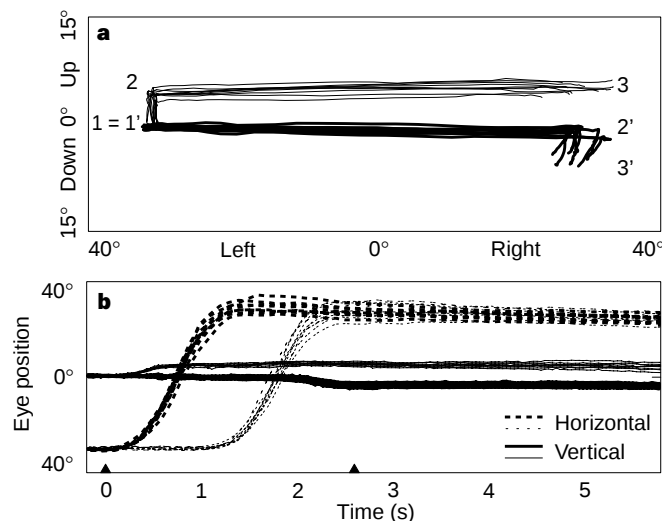


Figure 3 The actual VOR is non-commutative. A typical subject is put through the two sequences of rotations from Fig. 1. **a**, Plots of eye position in the head show near-ideal non-commutativity: final eye positions 3 and 3' differ by about 10° vertically. **b**, The same data plotted versus time. Triangles on the abscissa mark where the sequences of body rotations start and end. The heavy and light solid lines are separated by about 10° when the motions are over, showing that the final, vertical eye positions differed depending on the order of body rotations.

motor neurons differ, depending on the order of head rotations. There must be non-commutative processing between the inner ear and the motor endplate.

Various plausible mechanisms can be devised for the non-commutativity in the VOR, but all require non-commutative operators in the brain. One simple model, compatible with known physiology, differs from the model in Fig. 2 in just one essential way: it generates eye-position commands by integrating, not head velocity multiplied by -1 , but head velocity multiplied by eye-position information. This multiplicative interaction, described in the Methods, is non-commutative, and it suffices to give the eye-position command an appropriate memory of the order of head rotations. Simulations of this model are indistinguishable from the ideal trajectories plotted in Fig. 1c.

Further experiments are needed to test this model and to locate the non-commutative operators anatomically. One fundamental question is the role of gravity sensors and the associated otolith pathways. Our model implies that gravity sensors are not essential (though they may contribute) to the non-commutativity of the VOR, a claim that can be tested by performing the experiment from Fig. 1 in weightlessness. Another question is whether the non-commutative operators are wired into the brainstem circuits of the VOR or are located elsewhere. If the brainstem contains the necessary operators, then lesions elsewhere, for instance in the cerebellum, should not affect the non-commutativity of the VOR.

The VOR was chosen for this study because it is a relatively simple system where non-commutative computation is necessary for optimal function, as the thought experiment in Fig. 1 shows. However, by similar reasoning, one should also expect non-commutative operators in many other brain systems where they are needed, for example in systems dealing with rotary geometry, such as the circuits for head and limb control, auditory and visual localization, space constancy and mental rotation of objects^{11,14,15}.

Methods

Subjects and stimuli. Our five subjects were healthy, aged 29 to 37, and gave informed consent; three were naive as to the purpose of the experiment, though this did not affect the results. Subjects were strapped into a rotary chair, with the head fixed to the chair by a helmet and bitebar. They were asked to maintain fixation on the centre of a target disc 5 cm across, 2 m away. Then, in the dark,

they were put through one of four sequences of rotations—left then CCW, CCW then left, right then CW, or CW then right—each repeated a total of 10 times in randomly ordered blocks of 5. All rotations were about the centre of the subject's head. Horizontal rotations were 60° and lasted 1.5 s, torsional rotations were 10° and lasted 0.75 s. With a 0.35-s pause between turns, each sequence of two rotations lasted 2.6 s, and was followed by 4 s stationary in darkness. Starting positions were chosen so that all horizontal rotations had earth-vertical axes, and so did not change the subject's orientation relative to gravity. This arrangement shows that gravity sensors cannot remove the need for non-commutative computations in the VOR: to keep the eyes on target, any gravity-based estimates of the subject's torsional rotation must be combined with other estimates of the horizontal rotation^{25,26}, and this combination must be non-commutative, if the eyes are to end up in distinct positions as in Fig. 3.

Eye movements. We recorded the position of the left eye 100 times per second using search coils²⁷. For analysis, we omitted all trials where saccades (sudden gaze shifts) occurred during the head rotations. All data figures and simulations express eye positions as gaze vectors, showing the horizontal and vertical components of a vector of length one that points forward along the gaze line.

Simulations. Figure 2 shows the predictions of Raphan's vector-pulley model of the VOR¹⁹, with $k_o = 0.25$. Vertical excursions in the figure are small owing to the model's low torsional gain and errors in its vertical eye-position commands. These errors show up even in the rotation between positions 1 and 2 because the eye starts at 30° left and the model does not correctly handle the non-commutative interaction with the preceding rotation(s) that brought the eye there. The non-commutative model described in the text is simplified from the one in ref. 5: both models match Figs 1c and 3, but the former used a four-component neural representation of eye position and no pulleys, whereas the present one uses three components and the linear-plant model¹² of the eye muscles, which is a simple way to model pulleys^{12–14,21}. The full equations for this non-commutative VOR model are: $e = -h - h \times x$, $dx/dt = e$, $m = kx + re$, $dq/dt = (m - kq)/r$. Here k and r are related to the stiffness and viscosity of the orbital tissues^{6,24}. All other variables are three-component vectors: q is eye position, m is motor-neuron activity, x is the eye-position command in the brainstem, h is the head-velocity signal from the inner ear, e is a command from vestibular nuclei to motor neurons, and $\times$ is the vector cross product. The model is simple and biologically plausible, requiring only scalar integration, addition, subtraction and multiplication, all of which are believed to be in the repertoire of neurons²⁸. The formula for e implies that VOR circuits are cross-coupled, with vertical head-velocity signals influencing horizontal eye-position commands, and so on. Such cross-connections are also indicated by experiments in which the VOR was trained to produce eye movements about one axis in response to head motion about an orthogonal axis^{29,30}.

Received 2 December 1998; accepted 30 March 1999.

- Hamilton, W. R. *Lectures on Quaternions* (Hodges, Dublin, 1853).
- Schutz, B. *Geometrical Methods of Mathematical Physics* (Cambridge Univ. Press, Cambridge, 1980).
- McCarthy, J. M. *An Introduction to Theoretical Kinematics* (North-Holland, Amsterdam, 1990).
- Westheimer, G. Kinematics of the eye. *J. Opt. Soc. Am.* **47**, 967–974 (1957).
- Tweed, D. & Vilis, T. Implications of rotational kinematics for the oculomotor system in three dimensions. *J. Neurophysiol.* **58**, 832–849 (1987).
- Carpenter, R. H. S. *The Movements of the Eyes* (Pion, London, 1988).
- Tweed, D. & Vilis, T. Geometric relations of eye position and velocity vectors during saccades. *Vision Res.* **30**, 111–127 (1990).
- Crawford, J. D. & Vilis, T. Axes of eye rotation and Listing's law during rotation of the head. *J. Neurophysiol.* **65**, 407–423 (1991).
- Minken, A. W. H., van Opstal, A. J. & Van Gisbergen, J. A. M. Three-dimensional analysis of strongly curved saccades elicited by double-step stimuli. *Exp. Brain Res.* **93**, 521–533 (1993).
- Hestenes, D. Invariant body kinematics I. Saccadic and compensatory eye movements. *Neural Netw.* **7**, 65–77 (1994).
- Hestenes, D. Invariant body kinematics: II. Reaching and neurogeometry. *Neural Netw.* **7**, 79–88 (1994).
- Tweed, D., Misslisch, H. & Fetter, M. Testing models of the oculomotor velocity-to-position transformation. *J. Neurophysiol.* **72**, 1425–1429 (1994).
- Smith, M. A. & Crawford, J. D. Neural control of rotational kinematics within realistic vestibuloocular coordinate systems. *J. Neurophysiol.* **80**, 2295–2315 (1998).
- Tweed, D. A three-dimensional model of the human eye-head saccadic system. *J. Neurophysiol.* **77**, 654–666 (1997).
- Henriques, D. Y. P., Klier, E. M., Smith, M. A., Lowy, D. & Crawford, J. D. Gaze centered remapping of remembered visual space in an open-loop pointing task. *J. Neurosci.* **18**, 1583–1594 (1998).
- Van Opstal, A. J., Hepp, K., Hess, B. J. M., Straumann, D. & Henn, V. Two- rather than three-dimensional representation of saccades in monkey superior colliculus. *Science* **252**, 1313–1315 (1991).
- Schnabolk, C. & Raphan, T. Modeling three-dimensional velocity-to-position transformation in oculomotor control. *J. Neurophysiol.* **71**, 623–638 (1994).
- Straumann, D., Zee, D. S., Solomon, D., Lasker, A. G. & Roberts, D. Transient torsion during and after saccades. *Vision Res.* **35**, 3321–3334 (1995).

- Raphan, T. In *Three-dimensional Kinematics of Eye, Head and Limb Movements* (eds Fetter, M., Haslwanter, T., Misslisch, H. & Tweed, D.) 359–376 (Harwood, Amsterdam, 1997).
- Raphan, T. Modeling control of eye orientation in three dimensions. I. Role of muscle pulleys in determining saccadic trajectory. *J. Neurophysiol.* **79**, 2653–2667 (1998).
- Quaia, C. & Optican, L. M. Commutative saccadic generator is sufficient to control a 3-d ocular plant with pulleys. *J. Neurophysiol.* **79**, 3197–3215 (1998).
- Miller, J. M., Demer, J. L. & Rosenbaum, A. L. Effect of transposition surgery on rectus muscle paths by magnetic resonance imaging. *Ophthalmology* **100**, 475–487 (1993).
- Demer, J. L., Miller, J. M. & Poukens, V. Surgical implications of the rectus extraocular muscle pulleys. *J. Pediatr. Ophthalmol. Strabismus* **33**, 208–218 (1996).
- Robinson, D. A. Oculomotor unit behavior in the monkey. *J. Neurophysiol.* **35**, 393–404 (1970).
- Glaser, S. Interaction of semicircular canals and otoliths in the processing structure of the subjective zenith. *Ann. NY Acad. Sci.* **656**, 847–849 (1992).
- Merfeld, D. M. Modeling the vestibulo-ocular reflex of the squirrel monkey during eccentric rotation and roll tilt. *Exp. Brain Res.* **106**, 123–134 (1995).
- Robinson, D. A. A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans. Biomed. Eng.* **10**, 137–145 (1963).
- Koch, C. *Biophysics of Computation: Information Processing in Single Neurons* 471–473 (Oxford Univ. Press, New York, 1999).
- Schultheis, L. W. & Robinson, D. A. Directional plasticity of the vestibuloocular reflex in the cat. *Ann. NY Acad. Sci.* **374**, 504–512 (1981).
- Peng, G. C., Baker, J. F. & Peterson, B. W. Dynamics of directional plasticity in the human vertical vestibulo-ocular reflex. *J. Vestib. Res.* **4**, 453–460 (1994).

Acknowledgements. We thank D. M. Broussard, J. D. Crawford, J. Dichgans, C. E. Hawkins, J. A. Sharpe and T. Vilis for comments on the manuscript. This work was supported by the MRC of Canada and the Deutsche Forschungsgemeinschaft.

Correspondence and requests for materials should be addressed to D.B.T.

Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease

Victor O. Ona*, Mingwei Li*, Jean Paul G. Vonsattel†, L. John Andrews*, Sohail Q. Khan*, Woosik M. Chung‡, Ariel S. Frey‡, Anil S. Menon‡, Xiao-Jiang Li§, Philip E. Stieg*, Junying Yuan||, John B. Penney‡, Anne B. Young‡, Jang-Ho J. Cha‡ & Robert M. Friedlander*¶

* Neurosurgical Service, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA

† Department of Neuropathology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

‡ Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA

§ Department of Genetics, Emory University School of Medicine, Atlanta, Georgia 30322, USA

|| Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Huntington's disease is an autosomal-dominant progressive neurodegenerative disorder resulting in specific neuronal loss and dysfunction in the striatum and cortex¹. The disease is universally fatal, with a mean survival following onset of 15–20 years and, at present, there is no effective treatment. The mutation in patients with Huntington's disease is an expanded CAG/polyglutamine repeat in huntingtin, a protein of unknown function with a relative molecular mass of 350,000 (M_r 350K)². The length of the CAG/polyglutamine repeat is inversely correlated with the age of disease onset. The molecular pathways mediating the neuropathology of Huntington's disease are poorly understood. Transgenic mice expressing exon 1 of the human *huntingtin* gene with an expanded CAG/polyglutamine repeat develop a progressive syndrome with many of the characteristics of human Huntington's disease³. Here we demonstrate evidence of caspase-1 activation in the brains of mice and humans with the disease. In this transgenic mouse model of Huntington's disease, expression of a dominant-negative caspase-1 mutant extends survival and delays the appearance of neuronal inclusions, neurotransmitter receptor alterations and onset of symptoms, indicating that

caspase-1 is important in the pathogenesis of the disease. In addition, we demonstrate that intracerebroventricular administration of a caspase inhibitor delays disease progression and mortality in the mouse model of Huntington's disease.

Transgenic mice expressing exon 1 of human *huntingtin* with expanded CAG/polyglutamine repeat lengths under the control of its endogenous promoter develop a progressive, ultimately fatal neurological phenotype with many features of human Huntington's disease³. Of these transgenic mouse lines, the R6/2 line has been the most thoroughly studied^{3–5}. R6/2 mice develop normally, but by 5–7 weeks of age show significant loss of brain and body weight compared to control wild-type mice^{3,4}. Features of juvenile Huntington's disease, including irregular gait, abrupt shuddering movements, resting tremors and epileptic seizures, begin at about nine weeks of age. Ubiquitinated neuronal intranuclear inclusions (NII) can be found in striatal neurons of R6/2 mice by four weeks of age⁴. NII have also been found in the brains of human Huntington's disease patients⁶, but the role of NII in the pathogenesis of Huntington's disease is unknown.

The caspases are a family of cysteine proteases that are important for regulating apoptosis⁷. To evaluate a possible role for caspases in Huntington's disease, we crossbred R6/2 mice with a well characterized transgenic mouse expressing a dominant-negative mutant of caspase-1 in the brain⁸. In the mutant caspase-1 protein (M17Z), the active-site cysteine is substituted with a glycine. The neuron-specific enolase promoter targets the expression of mutant caspase-1 to neurons and glia within the central nervous system (NSE M17Z)^{9,10}. The M17Z protein can act as a dominant-negative inhibitor of caspase-1 pathways, as demonstrated by inhibiting lipopolysaccharide- and cerebral ischaemia-mediated pro-interleukin-1 β (pro-IL-1 β) processing^{8,11}. However, the mutant caspase-1 protein might also affect caspases other than caspase-1. The NSE M17Z mouse has normal embryonic development and brain size, and has been used as a tool to evaluate the role of caspase-1-mediated cell death in cerebral ischaemia and in amyotrophic lateral sclerosis (ALS)^{8,12}. Decreased infarct size following ischaemia and delayed progression of ALS have been demonstrated in the NSE M17Z mouse.

R6/2 and R6/2-NSE M17Z mice develop normally and are

indistinguishable from their wild-type littermates until about seven weeks of age. Progression of motor dysfunction was objectively evaluated using a RotaRod. R6/2-NSE M17Z mice performed significantly better than R6/2 mice throughout the test period (Fig. 1a). The deterioration of motor coordination seen in R6/2 mice is also seen in R6/2-NSE M17Z mice at a later stage (Fig. 1b). These results show that expression of the caspase-1 mutant delays the progression of motor dysfunction in R6/2 mice. To verify that the differences were not due to different levels of mutant *huntingtin* transgene expression in R6/2 and R6/2-NSE M17Z mice, we performed quantitative *in situ* hybridization. Equal expression was detected in R6/2 and R6/2-NSE M17Z mice. There was no transgene expression in either wild-type or NSE M17Z mice (data not shown).

Huntington's disease patients show significant weight loss despite adequate caloric intake, indicating a hypermetabolic state¹. The aetiology of weight loss and hypermetabolism are unknown, and they do not appear to result solely from hyperkinesia. Similarly, R6/2 mice gain weight normally up to 6–7 weeks of age, and thereafter progressively lose weight to about 60% of that of controls³. The weight curve of R6/2-NSE M17Z mice is intermediate between those of control and R6/2 mice. In addition, the weight of the R6/2-NSE M17Z mice does not decrease to the same level before death as in their R6/2 counterparts (12.4 g versus 8.8 g, $P < 0.05$) (Fig. 1c). NSE-M17Z mice weighed the same as wild-type mice. As the NSE promoter targets transgene expression to the brain and testis, these results indicate that weight loss in R6/2 mice results partly from a systemic manifestation of disease progression within the nervous system⁹. However, as mutant huntingtin is expressed in peripheral tissues, low levels of M17Z expression might affect a putative role for peripheral huntingtin as a mediator of weight loss³.

Elevation of mature IL-1 β levels has been used as sensitive and specific evidence for caspase-1 activation, as caspase-1 is the main (if not the only) enzyme to activate pro-IL-1 β *in vivo*¹³. To evaluate caspase-1 activation in R6/2 mice, we measured mature IL-1 β in brain lysate of 12-week-old symptomatic mice. Levels of mature IL-1 β in R6/2 mice were elevated to 268% of those in wild-type controls ($P = 0.007$). The increase in mature IL-1 β was signifi-

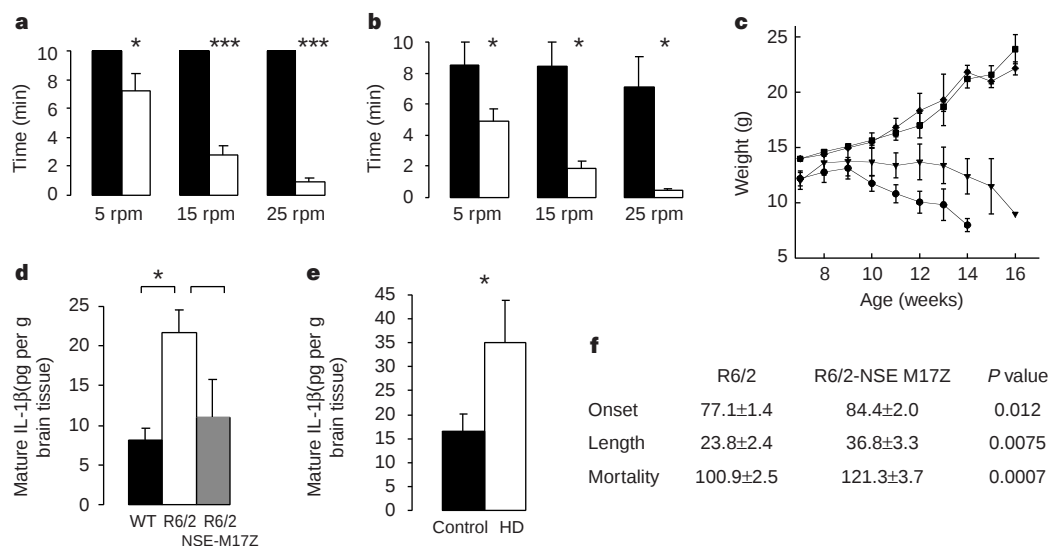


Figure 1 Progression of disease in R6/2 mice is mediated by a caspase-1 dominant-negative-sensitive pathway. **a**, Nine-, and **b**, twelve-week-old mice were tested on a RotaRod (Columbus Instruments) at three speeds. Testing was terminated either when the mouse fell from the rod or at 10 min if it remained on the rod. Black bars, R6/2-NSE M17Z; white bars, R6/2 mice. rpm, rotations per minute. **c**, Body weight. Female mice of the indicated genotypes were used to generate weight curves ($n = 6$ –8 per group). Squares, wild type; diamonds, NSE

M17Z; circles, R6/2; triangles, R6/2-NSE M17Z. **d**, Mature IL-1 β levels in 12-week-old mice ($n = 4$). **e**, Mature IL-1 β levels in grade 3–4 patients (HD) ($n = 5$) and control brains ($n = 6$). **f**, Disease onset, length and mortality in days are recorded as mean $\pm$ s.e.m. Complete consecutive litters of the same generation were used for this part of the study. R6/2, $n = 14$; R6/2-NSE M17Z, $n = 7$. Error bars indicate s.e.m. Asterisk, $P < 0.05$; three asterisks, $P < 0.0005$.

cantly inhibited in 12-week-old R6/2-NSE M17Z mice (Fig. 1d). Mature IL-1 β was not raised in eight-week-old asymptomatic R6/2 mice. Thus, the increase in mature IL-1 β in symptomatic R6/2 mice is inhibited in caspase-1 dominant-negative mutants. To evaluate the relevance of this finding, we measured mature IL-1 β levels in brains of human patients. In human grade 3–4 Huntington's disease cortex, mature IL-1 β was significantly increased to 213% of that in normal controls ($P < 0.05$) (Fig. 1e).

To evaluate the impact of caspase-1 dominant-negative inhibition on Huntington's disease, we monitored disease progression in R6/2 and R6/2-NSE M17Z mice. Disease onset was scored as the first date when coarse coating or limb tremors were detected. Mice were scored three times per week by two trained observers blinded to the genotype and age of the mice. Disease onset and mortality were significantly delayed in the R6/2-NSE M17Z mice by 7.3 days and 20.4 days, respectively, when compared to R6/2 littermates (Fig. 1f). The protection conferred by M17Z expression represents a 55% increase in disease duration and a 20% prolongation of life. In addition, to rule out a strain-related epigenetic effect mediating protection, we treated seven-week-old R6/2 mice with the caspase inhibitor zVAD-fmk by continuous intracerebroventricular infusion for four weeks. zVAD-fmk-treated mice performed better on RotaRod (Fig. 2a) and lived 25% longer (98.6 versus 79.0 days, $n = 5$ /group, $P = 0.002$) (Fig. 2b) than control mice treated with zFA-fmk.

There is controversy about the role of NII in Huntington's disease^{4,6,14}. There is a close temporal relationship between NII detection and disease onset^{4,6}. The appearance of NII is significantly delayed in R6/2-NSE M17Z mice when compared to R6/2 littermates. Nine-week-old R6/2 mice showed abnormal motor symptoms and abundant NII, as determined by ubiquitin and amino-terminal huntingtin immunohistochemistry (Fig. 3a, c), whereas R6/2-NSE M17Z mice were asymptomatic at nine weeks, and none had NII (Fig. 3b, d). NII were detected in twelve-week-old symptomatic R6/2-NSE M17Z mice (Fig. 3f). These results were replicated using two separate N-terminal huntingtin antibodies. Reactive fibrillary astrogliosis has been described as a significant finding in the brains of human patients. Reactive astrogliosis was detected in the cortex and striatum of nine-week-old R6/2 mice, as graded by a blinded neuropathologist (J.P.V.). Slight reactive astrogliosis was detected in one out of four age-matched R6/2-NSE M17Z mice, and none was detected in the other three R6/2-NSE M17Z mice or in the wild-type or NSE M17Z mice ($n = 4$ per genotype) (Fig. 3g, h).

There is an age-dependent decrease in neurotransmitter receptor binding and messenger RNA levels in R6/2 mice⁵. This decrease in

adenosine A2a, dopamine D1 and dopamine D2 receptor binding and dopamine D2 receptor mRNA in R6/2 mice is significantly inhibited in R6/2-NSE M17Z mice (Fig. 4a). Functional deterioration in these mice correlated with the decrease in neurotransmitter density⁵, so the improved functional status of the R6/2-NSE M17Z mice compared with the R6/2 mice is probably due in part to the delay in neurotransmitter-receptor loss. The densities of adenosine A2a, dopamine D1 and dopamine D2 receptors were not statistically significantly different between NSE M17Z and wild-type littermate mice.

Caspase-1 cleaves mutant and wild-type huntingtin *in vitro*¹⁵. To investigate the mechanism underlying the protection conferred by mutant caspase-1, we evaluated cleavage of endogenous huntingtin using a well characterized carboxy-terminal huntingtin antibody¹⁶. Compared to wild-type and NSE-M17Z mice, nine-week-old R6/2 mice show differential processing of huntingtin, with a prominent cleavage product revealed by western blotting. R6/2-NSE M17Z mice, on the other hand, do not show the cleavage band seen in R6/2 mice (Fig. 4b). Immunohistochemistry for endogenous huntingtin¹⁷ showed diminished staining in the hippocampus, striatum, cortex and cerebellum of R6/2 mice as compared to wild-type and R6/2-NSE M17Z mice (Fig. 4c–k). The specific role of huntingtin cleavage fragments in Huntington's disease and in the formation of NII remains controversial. However, as cleavage of

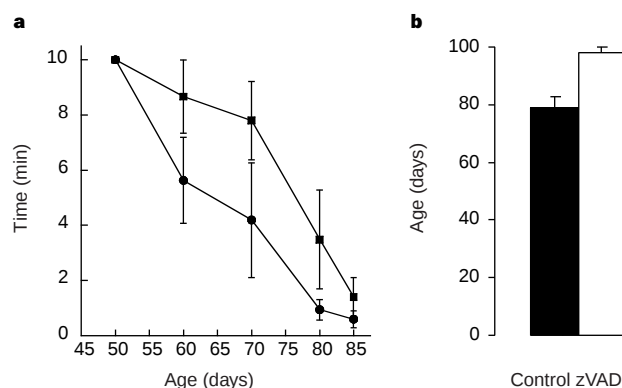


Figure 2 zVAD-fmk delays progression and mortality in R6/2 mice. **a**, Control zFA-fmk (circles) and zVAD-fmk (squares) treated R6/2 mice were tested on the RotaRod (15 r.p.m.), beginning the day before the placement of osmotic pumps for intracerebroventricular drug administration. **b**, Intracerebroventricular administration of zVAD-fmk delays mortality in R6/2 mice ($n = 5$ per group, $P < 0.0002$).

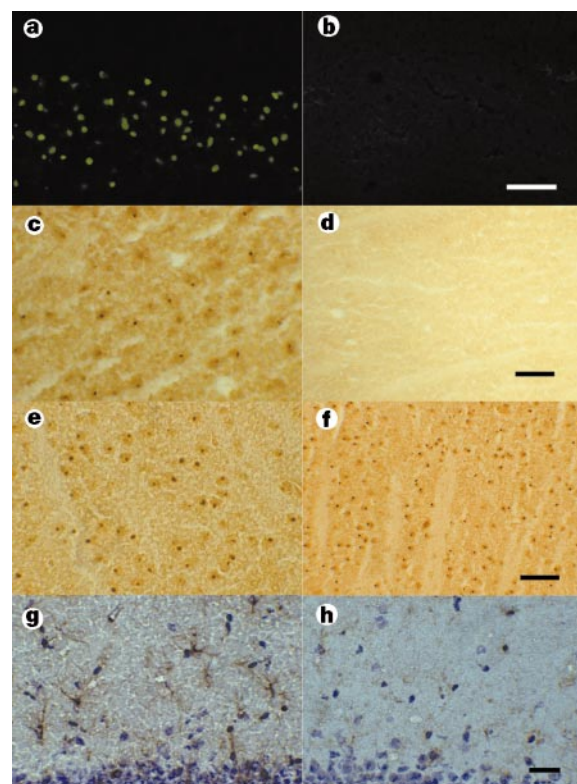


Figure 3 NII and astrogliosis are inhibited in R6/2-NSE M17Z mice. **a, b**, Fluorescent immunohistochemistry demonstrating ubiquitin-positive inclusions in the CA1 region of hippocampus in 9-week-old R6/2 mice (**a**), but not in R6/2-NSE M17Z mice (**b**). Scale bar, 50 μ m. **c, d**, Peroxidase immunohistochemistry with EM48, an antibody directed against the N-terminal region of huntingtin, demonstrates NII in the striatum of 9-week-old R6/2 mice (**c**) but not of R6/2-NSE M17Z mice (**d**). Scale bar, 40 μ m. **e, f**, At 12 weeks of age, ubiquitin-positive NII are present in the striatum of both R6/2 (**e**) and R6/2-NSE M17Z (**f**) mice, although the inclusions in the R6/2-NSE M17Z mice are smaller. Scale bar, 40 μ m. **g, h**, Glial fibrillary acidic protein (GFAP) immunohistochemistry (brown peroxidase product with blue haematoxylin counterstain) in the hippocampal dentate region of 9-week-old R6/2 mice (**g**), and R6/2-NSE M17Z mice (**h**). Scale bar, 40 μ m.

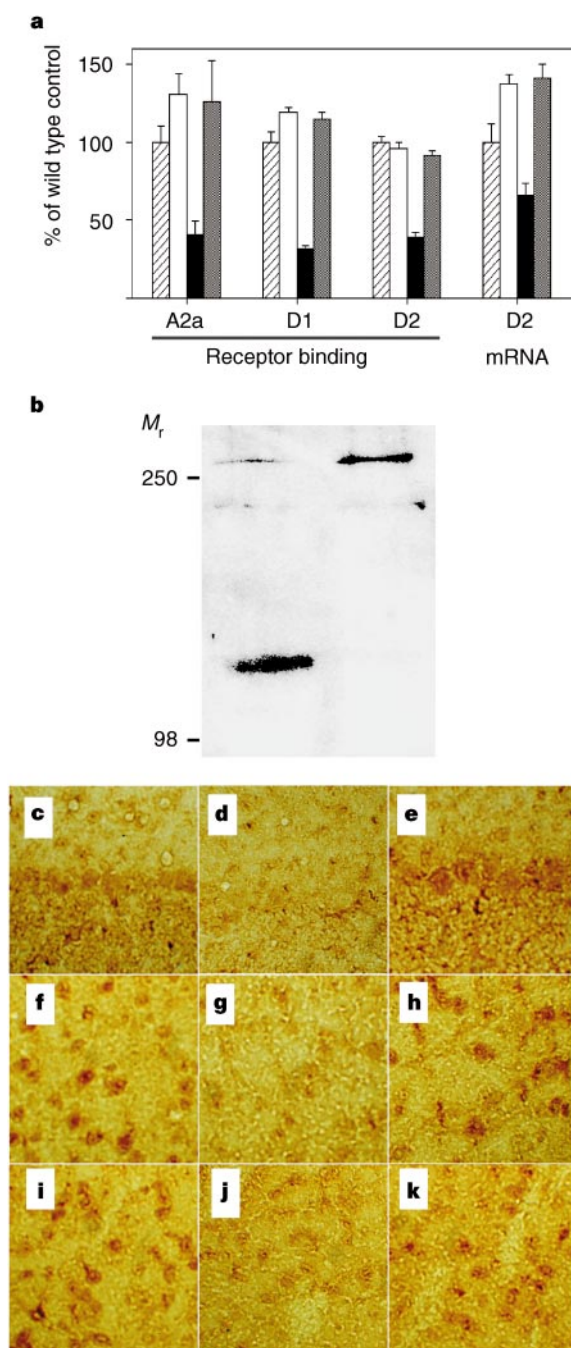


Figure 4 NSE M17Z inhibits neurotransmitter-receptor loss and cleavage of endogenous huntingtin in R6/2 mice. **a**, R6/2 mice have specific decreases in neurotransmitter-receptor binding preceded by decreases in corresponding mRNA species⁵. At nine weeks of age, R6/2 mice (black bars) had significantly less receptor binding than wild-type (hatched bars), NSE M17Z (white bars) or R6/2 NSE M17Z (shaded bars) mice. In addition, R6/2 mice had significantly less dopamine-D2-receptor mRNA than other groups of mice (analysis of variance, $P < 0.05$ for all assays). Baseline differences between wild-type and NSE-M17Z receptor levels were not statistically significant. **b**, Western-blot analysis using an antibody against amino acids 1,981–2,580 of human huntingtin¹⁶ shows differential processing of endogenous huntingtin in R6/2 (left lane) and R6/2-NSE M17Z (right lane) mice. Relative molecular mass markers are shown on the left (in thousands). **c–k**, Peroxidase immunohistochemistry using the same anti-huntingtin antibody¹⁷ in the Purkinje-cell layer of the cerebellum (**c**, **d**, **e**), frontal cortex (**f**, **g**, **h**) and striatum (**i**, **j**, **k**) in wild-type (**c**, **f**, **i**), R6/2 (**d**, **g**, **j**) and R6/2-NSE M17Z (**e**, **h**, **k**) mice. R6/2 mice consistently demonstrate reduced immunoreactivity for endogenous huntingtin, as compared to wild-type and R6/2-NSE M17Z mice.

wild-type and mutant huntingtin *in vitro* by caspase-1 has been demonstrated, the above evidence implicates caspase-1 in the generation of wild-type and mutant huntingtin fragments *in vivo*¹⁵. Apoptotic features have been described in striatal cells transfected with wild-type full-length huntingtin¹⁸.

We demonstrate that motor weakness, NII formation, neurotransmitter receptor alteration and, most important, symptomatic onset and mortality in R6/2 mice is delayed by expression of a dominant-negative caspase-1 mutant. In addition, we have delayed mortality of R6/2 mice by chronic intracerebroventricular administration of a caspase inhibitor. We propose that caspases inhibitable by mutant caspase-1 are important in the Huntington's disease-like phenotype of the R6/2 transgenic mouse. Elevation of mature IL-1 β levels in symptomatic R6/2 mice, which was inhibited in R6/2-NSE M17Z mice, indicates that caspase-1 may be critical for the symptomatic disease progression of R6/2 mice. In addition, detection of increased mature IL-1 β in the brains of Huntington's disease patients adds significance to these findings. Reductions in specific neuronal populations have been identified in R6/2 mice of more than 12 weeks old (S. Davies, personal communication). However, neuronal dysfunction appears to be the dominant factor mediating the appearance of symptoms and death in the R6/2 mouse^{3,5}; lack of significant cell loss is probably secondary to rapid disease progression, which does not allow significant cell death to occur³. Evidence for apoptosis has been seen in Huntington's disease patients and *in vitro* using transfection with mutant huntingtin^{14,19,20}. In addition, a transgenic mouse expressing full-length huntingtin shows significant apoptotic neurodegeneration and behavioural characteristics of Huntington's disease²¹. As the expression of this dominant-negative caspase-1 mutant delays neuronal dysfunction as well as death, our data implicate caspase-1-mediated pathways in the symptomatic progression of R6/2 mice. Although R6/2 mice have diabetes²², which probably affects body weight, the presence of NII and alterations in neurotransmitter receptors are unlikely to be related to diabetic effects.

In acute cellular injury, such as ischaemia, caspases are activated sufficiently to mediate cell death^{8,11,23}. In a chronic neurodegenerative disease, certain caspases may be activated at a sub-lethal threshold, insufficient to cause cell death but enough to cause cellular abnormalities, manifested in R6/2 mice as neuronal dysfunction. These results indicate a novel role for caspase-1 not only in cell death, but also in mediating cellular dysfunction leading to symptomatic progression and death. Caspase-1 and caspase-3 (but not caspase-7 or caspase-8) cleave huntingtin, and caspase-1 can activate caspase-3 *in vitro*^{15,24,25}. Given that endogenous murine huntingtin is cleaved in R6/2 symptomatic mice, one possible mechanism is that the mutant exon 1 of huntingtin initiates a toxic state within susceptible neuronal populations, leading to caspase-1 activation and cleavage of certain proteins, including huntingtin. Caspase-1 could also activate caspase-3, mediating further cleavage of endogenous huntingtin and increasing the concentration of small huntingtin fragments, accelerating the formation of NII.

This study provides *in vivo* functional mechanistic evidence of caspase-1 activation mediated by exon 1 of mutant huntingtin. Caspase-1 mediates pro-IL-1 β processing, generating secreted mature IL-1 β (ref. 26). Endogenously produced mature IL-1 β is involved in neuronal cell death^{27,28}, and reactive astrogliosis, which occurs as a response to cellular injury, is stimulated by IL-1 β ²⁹. The precise relationship of IL-1 β and astrogliosis to the progression of Huntington's disease is still unclear. Caspase-1-mediated pathways appear to be shared in a variety of acute and chronic neurological disorders (ischaemia, ALS, Huntington's disease and Parkinson's disease) as a common mechanism mediating cell dysfunction and cell death^{8,12,23,30}. Our findings show that caspase-1 plays an important and specific role in the early stages of Huntington's disease, as demonstrated by the inhibition of both NII formation and the specific neuroreceptor abnormalities found in the disease. Caspase-1

inhibition delays onset and prolongs survival in a transgenic mouse model of Huntington's disease, and is potentially applicable to the human disease. □

Methods

Animals. Four male NSE M17 mice were crossed with eight wild-type female recipients of transplanted ovaries from R6/2 mice (Jackson Laboratories). Polymerase chain reaction was used for genotyping^{3,8}. We used complete consecutive litters up to a certain birth date for symptomatic studies. Litters generated thereafter were used for the remaining studies.

Mature IL-1 β determination. Mature IL-1 β quantification was done as previously described using an ELISA kit (Genzyme)²³.

Intracerebroventricular implantation of osmotic pumps. Spontaneously breathing seven-week-old R6/2 mice were anaesthetized, and a 2-cm sagittal incision was made over the skull. A small opening was made on the cranium 1 mm right lateral to the bregma. A microcannula was inserted to a depth of 3 mm, secured with a dental acrylic (Duralay, Bioanalytical Systems), and attached to an osmotic pump (Alzet). The pump was prefilled with 100 μ g per 20 g body weight of either zVAD-fmk or zFA-fmk (Enzyme Systems). The pumps continuously deliver the drug for four weeks. The experimenter was blinded to the identity of the drug used in the pump until the death of all the mice.

In situ hybridization. We constructed oligonucleotide probes targeted against a human-specific intron of the human Huntington's disease transgene sequence (sequence 5'-TCT GGG TTG CTG GGT CAC TCT GTC TCT GCG GAG CCG GGG G-3') and against the dopamine-D2-receptor gene (sequence 5'-GGC AGG GTT GGC AAT GAT ACA CTC ATT CTG GTC TGT ATT-3'). Probes were 3'-end labelled with ³⁵S-dATP using an oligonucleotide-labelling kit (NEN Life Science Products). Parasagittal sections (12 μ m) of frozen mouse brains were thaw-mounted onto poly-L-lysine-coated glass slides. Slides were fixed in 4% paraformaldehyde/0.1 M phosphate buffered-saline (PBS) for 10 min, washed in PBS for 3 $\times$ 5 min, acetylated with 0.25% acetic anhydride in 0.1 M triethanolamine, pH 8.0, for 10 min, washed in PBS for 5 min, and dehydrated in graded ethanol solutions. Slides were incubated overnight at 37 °C with 80 μ l of hybridization buffer (50% formamide, 0.3 M NaCl, 10 mM Tris, 1 mM EDTA, 10% dextran sulphate, 1 $\times$ Denhart's solution) containing radioactively labelled probe (30,000 c.p.m. per μ l). After hybridization, slides were washed in a series of sodium citrate-sodium chloride washes, rinsed with 70% ethanol and air-dried. The slides were apposed to β -max film (Amersham) and exposed for 2 weeks before the film was developed. Film densities were analysed using a computer-based image-analysis system (M1, Imaging Research).

Neuronal intranuclear inclusion and neurotransmitter receptor analysis. Receptor binding was done in parallel on fresh frozen brain tissue from wild-type, NSE M17Z, R6/2 and R6/2-NSE M17Z mice ($n = 4$ animals per group) as described⁵. Immunohistochemistry was performed on 12- μ m brain sections using antibodies against ubiquitin (Dako), glial fibrillary acidic protein or huntingtin¹⁶ as described¹⁷. Western blotting was done according to standard techniques using an anti-huntingtin antibody (HF-1; courtesy of M. MacDonald).

Received 22 February; accepted 14 April 1999.

- Harper, P. S. *Huntington's Disease* (Saunders, London, 1991).
- Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is unstable on Huntington's disease chromosomes. *Cell* **72**, 971–983 (1993).
- Mangiarini, L. *et al.* Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**, 493–506 (1996).
- Davies, S. W. *et al.* Formation of neuronal intranuclear inclusions underlies the neurological dysfunction in mice transgenic for the HD mutation. *Cell* **90**, 537–548 (1997).
- Cha, J. H. *et al.* Altered brain neurotransmitter receptors in transgenic mice expressing a portion of an abnormal human huntington disease gene. *Proc. Natl Acad. Sci. USA* **95**, 6480–6485 (1998).
- DiFiglia, M. *et al.* Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* **277**, 1990–1993 (1997).
- Alnemri, E. S. *et al.* Human ICE/CED-3 protease nomenclature. *Cell* **87**, 171 (1996).
- Friedlander, R. M. *et al.* Expression of a dominant negative mutant of interleukin-1 beta converting enzyme in transgenic mice prevents neuronal cell death induced by trophic factor withdrawal and ischemic brain injury. *J. Exp. Med.* **185**, 933–940 (1997).
- Forss-Petter, S. *et al.* Transgenic mice expressing beta-galactosidase in mature neurons under neuron-specific enolase promoter control. *Neuron* **5**, 187–197 (1996).
- Burne, J. F., Staple, J. K. & Raff, M. C. Glial cells are increased proportionally in transgenic optic nerves with increased numbers of axons. *J. Neurosci.* **16**, 2064–2073 (1996).
- Hara, H. *et al.* Attenuation of transient focal cerebral ischemic injury in transgenic mice expressing a mutant ICE inhibitory protein. *J. Cereb. Blood Flow Metab.* **17**, 370–375 (1997).
- Friedlander, R. M., Brown, R. H., Gagliardini, V., Wang, J. & Yuan, J. Inhibition of ICE slows ALS in mice. *Nature* **388**, 31 (1997).
- Li, P. *et al.* Mice deficient in IL-1 beta-converting enzyme are defective in production of mature IL-1 beta and resistant to endotoxic shock. *Cell* **80**, 401–411 (1995).

- Saudou, E., Finkbeiner, S., Devys, D. & Greenberg, M. E. Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell* **95**, 55–66 (1998).
- Wellington, C. L. *et al.* Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *J. Biol. Chem.* **273**, 9158–9167 (1998).
- Persichetti, F. *et al.* Differential expression of normal and mutant Huntington's disease gene alleles. *Neurobiol. Dis.* **3**, 183–190 (1996).
- Kosinski, C. M. *et al.* Huntingtin immunoreactivity in the rat neostriatum: differential accumulation in the projection and interneurons. *Exp. Neurol.* **144**, 239–247 (1997).
- Kim, M. *et al.* Mutant huntingtin expression in clonal striatal cells: dissociation of inclusion formation and neuronal survival by caspase inhibition. *J. Neurosci.* **19**, 964–973 (1999).
- Portera-Cailliau, C., Hedreen, J. C., Price, D. L. & Koliatos, V. E. Evidence for apoptotic cell death in Huntington disease and excitotoxic animal models. *J. Neurosci.* **15**, 3775–3787 (1995).
- Lunkes, A. & Mandel, J. L. A cellular model that recapitulates major pathogenic steps of Huntington's disease. *Hum. Mol. Genet.* **7**, 1355–1361 (1998).
- Reddy, P. H. *et al.* Behavioural abnormalities and selective neuronal loss in HD transgenic mice expressing mutated full-length HD cDNA. *Nature Genet.* **20**, 198–202 (1998).
- Hurlbert, M. S. *et al.* Mice transgenic for an expanded CAG repeat in Huntington's disease gene develop diabetes. *Diabetes* **8**, 649–651 (1999).
- Hara, H. *et al.* Inhibition of interleukin 1 beta converting enzyme family proteases reduces ischemic and excitotoxic neuronal damage. *Proc. Natl Acad. Sci. USA* **94**, 2007–2012 (1997).
- Goldberg, Y. P. *et al.* Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nature Genet.* **13**, 442–449 (1996).
- Tewari, M. *et al.* Yama/CPP32 beta, a mammalian homolog of CED-3, is a CrmA-inhibitable protease that cleaves the death substrate poly(ADP-ribose) polymerase. *Cell* **81**, 801–809 (1995).
- Thornberry, N. A. *et al.* A novel heterodimeric cysteine protease is required for interleukin-1 beta processing in monocytes. *Nature* **356**, 768–774 (1992).
- Troy, C. M., Stefanis, L., Prochiantz, A., Greene, L. A. & Shelanski, M. L. The contrasting roles of ICE family proteases and interleukin-1 beta in apoptosis induced by trophic factor withdrawal and by copper/zinc superoxide dismutase down-regulation. *Proc. Natl Acad. Sci. USA* **93**, 5635–5640 (1996).
- Friedlander, R. M., Gagliardini, V., Rotello, R. J. & Yuan, J. Functional role of interleukin 1 beta (IL-1 beta) in IL-1 beta-converting enzyme-mediated apoptosis. *J. Exp. Med.* **184**, 717–724 (1996).
- Lee, S. C., Dickson, D. W. & Brosnan, C. F. Interleukin-1, nitric oxide and reactive astrocytes. *Brain Behav. Immun.* **9**, 345–354 (1995).
- Klivenyi, P., Andreassen, O., Friedlander, R. M. & Beal, M. F. Transgenic mice expressing a dominant negative mutant interleukin-1 beta converting enzyme shows resistance to MPTP neurotoxicity. *NeuroReport* **10**, 635–638 (1999).

Acknowledgements. We thank E. Signer for advice and support; C. Swap, A. Dunah and W. Hobbs for expert technical assistance; M. MacDonald for providing Huntington antibodies; and E. Balodimas for editorial assistance. This work was supported by a grant from the Hereditary Disease Foundation (to R.M.F. and J.H.J.C.), the Huntington's Disease Society of America (to J.-H.J.C.), the Glendon Foundation (to A.B.Y.) and by grants from the NIH (to J.-H.J.C., A.B.Y. and J.Y.). This paper is dedicated to J.B.P., who died during the review of this manuscript.

Correspondence and requests for materials should be addressed to R.M.F. (e-mail: rfriedlander@rics.bwh.harvard.edu).

In vivo cell sorting in complementary segmental domains mediated by Eph receptors and ephrins

Qiling Xu, Georg Mellitzer, Vicky Robinson & David G. Wilkinson

Division of Developmental Neurobiology, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

The restriction of intermingling between specific cell populations is crucial for the maintenance of organized patterns during development. A striking example is the restriction of cell mixing between segments in the insect epidermis¹ and the vertebrate hindbrain² that may enable each segment to maintain a distinct identity. In the hindbrain, this is a result of different adhesive properties of odd- and even-numbered segments (rhombomeres)^{3,4}, but an adhesion molecule with alternating segmental expression has not been found. However, blocking experiments suggest that Eph-receptor tyrosine kinases may be required for the segmental restriction of cells⁵. Eph receptors and their membrane-bound ligands, ephrins, are expressed in complementary rhombomeres⁶ and, by analogy with their roles in axon pathfinding^{7,8}, could mediate cell repulsion at boundaries. Remarkably, transmembrane ephrins can themselves transduce signals^{9,10}, raising the possibility that bi-directional signalling occurs between adjacent ephrin- and Eph-receptor-expressing

cells. We report here that mosaic activation of Eph receptors leads to sorting of cells to boundaries in odd-numbered rhombomeres, whereas mosaic activation of ephrins results in sorting to boundaries in even-numbered rhombomeres. These data implicate Eph receptors and ephrins in the segmental restriction of cell intermingling.

Interactions between Eph receptors and ephrins largely fall into two classes: EphA receptors bind ephrin-A ligands anchored with glycosyl phosphatidylinositol, whereas EphB receptors bind the transmembrane ephrin-B proteins; an exception is EphA4, which binds ephrin-B2 and -B3 as well as ephrin-A ligands¹¹ (see ref. 12 for nomenclature). EphA4, EphB2 and EphB3 are expressed in rhombomeres r3/r5, whereas ephrin-B1, -B2 and -B3 are expressed in r2/r4/r6 (ref. 6). As a result of this complementary expression, interactions of EphA4 and EphB receptors with ephrin-B proteins will occur at the interface of adjacent rhombomeres. The finding that expression of a truncated EphA4 receptor leads to the presence of r3/r5 cells in r2/r4/r6 territory suggested that this receptor might regulate the identity of cells, or restrict intermingling between segments⁵. To distinguish between these possibilities, we analysed whether mosaic activation of EphA4 and EphB receptors leads to changes in the identity or movement of cells within r3/r5.

Ephrin-B2 was expressed in a mosaic fashion by injecting *ephrin-B2* messenger RNA into one cell of eight-cell-stage zebrafish embryos, together with *lacZ* mRNA in order to detect the expressing cells by staining for β -galactosidase activity. As a consequence, endogenous EphA4 and EphB receptors will be activated in r3/r5

cells that are in contact with cells ectopically expressing ephrin-B2 (Fig. 1a). In control injections of *lacZ* mRNA alone, the expressing cells have a scattered distribution because of the extensive mixing of cells during early zebrafish development (Fig. 1c). In contrast, cells expressing ephrin-B2 become restricted to the boundaries of r3/5, whereas in r2/r4/r6 many expressing cells are present in central regions of the segment (Fig. 1d, e). The expression patterns of *krox-20* or *EphA4*, markers of r3/r5 identity, are not altered, indicating that the mosaic expression of ephrin-B2 does not alter the identity of the expressing or adjacent cells.

Because full-length ephrin-B2 was expressed in these experiments, it was possible that signal transduction through this ephrin was required for the segregation of cells in r3/r5. To examine this, we expressed truncated ephrin-B2 in a mosaic fashion. This truncated ephrin lacks the cytoplasmic region, including conserved tyrosine residues that are phosphorylated upon binding to Eph receptor^{9,10} (Fig. 1b), but activates EphA4 phosphorylation in adjacent cells (Fig. 2a). We found a segregation of expressing cells adjacent to the boundaries of r3/r5, but a scattered distribution in r2/r4/r6 (Fig. 1f, g), similar to the results with full-length ephrin-B2. These data indicate that mosaic activation of Eph receptors by ephrin-B2 is sufficient for cell sorting, but they do not address the possibility that signal transduction through ephrin-B proteins could, in the absence of Eph receptor activation, also mediate cell sorting.

To determine whether signal transduction through ephrin-B proteins can trigger a response, we analysed the distribution of

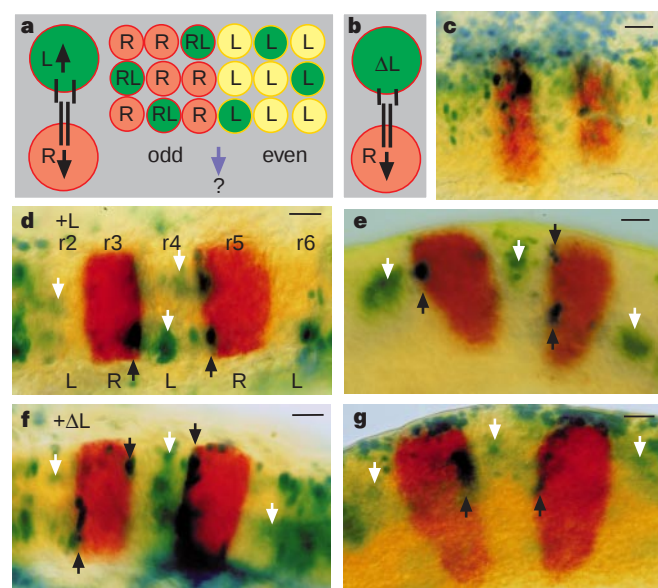


Figure 1 Cell sorting after mosaic expression of ephrin-B2. **a**, Left side; diagram illustrating activation of EphA4 and EphB receptors (R) by ephrin-B2 ligand (L). This interaction also leads to activation of ephrin-B2. Right side: following co-injection of *lacZ* and *ephrin-B2* RNA into one cell at the eight-cell stage, there will initially be a random distribution of cells co-expressing β -galactosidase and ephrin-B2 (green) in presumptive r3/r5 (orange) and r2/r4/r6 (yellow) territory. This will lead to mosaic activation of EphA4 and EphB receptors in r3/r5. **b**, Diagram illustrating activation of EphA4 and EphB receptors by truncated ephrin-B2 (Δ L), which cannot itself transduce signals. **c-g**, After injection of various RNAs at the eight-cell stage, 17-h embryos were fixed, stained for β -galactosidase activity (blue-green signal), then *in situ* hybridization was carried out to detect *krox-20*, a marker of r3/r5 (orange-red signal). **c**, Expression of *lacZ* alone: β -galactosidase-expressing cells have a scattered distribution. **d, e**, Co-injection of *lacZ* and *ephrin-B2* RNA: expressing cells have a scattered distribution in r2/r4/r6 (white arrows indicate examples), but sort to the boundaries of r3/r5 (black arrows). **f, g**, Co-injection of *lacZ* and truncated *ephrin-B2* RNA: a similar sorting of expressing cells is observed. Scale bars, 50 μ m.

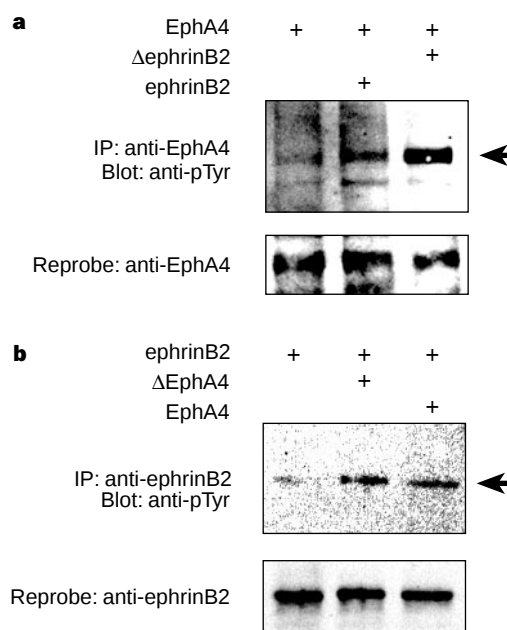


Figure 2 Activation of EphA4 and ephrin-B2 by truncated ligands. **a**, Induction of tyrosine phosphorylation of EphA4 by full-length and truncated ephrin-B2. Cells expressing EphA4 were mixed and incubated with uninjected cells or cells expressing ephrin-B2 or truncated ephrin-B2 (Δ ephrinB2). EphA4 was immunoprecipitated from cell lysates and phosphotyrosine detected on a western blot (top panel). The arrow indicates EphA4 protein. The blot was reprobed with anti-EphA4 antibody (lower panel). Full-length and truncated ephrin-B2 induce tyrosine phosphorylation of EphA4; the higher level of phosphorylation with truncated ephrin-B2 may reflect a greater stability of this protein. **b**, Induction of tyrosine phosphorylation of ephrin-B2 by full-length and truncated EphA4. Cells expressing ephrin-B2 were mixed with uninjected cells or cells expressing EphA4 or truncated EphA4 (Δ EphA4). Ephrin-B2 was immunoprecipitated from cell lysates, and tyrosine phosphorylation detected on a western blot (top panel). The arrow indicates ephrin-B2. The blot was reprobed with anti-ephrin-B2 antibody (lower panel). Full-length and truncated EphA4 induce tyrosine phosphorylation of ephrin-B2.

cells expressing truncated EphA4, which can bind to and activate ephrin-B proteins (Fig. 2b), but lacks the tyrosine kinase domain and juxtamembrane tyrosine residues implicated in signal transduction¹³ (Fig. 3a). Cells expressing truncated EphA4 were found to segregate adjacent to the boundaries of r2/r4/r6, whereas labelled cells are frequently present in central regions of r3/r5 (Fig. 3c–f). Cells expressing truncated EphB2 protein have a similar distribution (data not shown). Quantification of the distribution of labelled cells within rhombomeres confirms that cells expressing truncated EphA4 receptor sort towards boundaries in even-numbered segments, whereas cells expressing full-length or truncated ephrin-B2 sort towards boundaries in the complementary odd-numbered segments (Fig. 4). In some embryos injected with RNA encoding truncated EphA4, rhombomeres r3/r5 were altered in shape (data not shown), as previously found after injection at the two-cell stage⁵. However, this phenotype occurred in only 5% of the embryos after injection at the eight-cell stage compared with >50% after injection at the two-cell stage. These data are consistent with the scenario that dominant-negative blocking of endogenous EphA4 has increased the intermingling of r3/r5 cells into r2/r4/r6, but that local signals switch isolated ectopic cells to an even-numbered identity, whereas groups of cells can maintain their identity⁵.

The above analyses were carried out in embryos at the 12–15-somite stage of development, whereas presumptive r3/r5 marked by Krox-20 expression are first detected at the two-somite stage. Properties that restrict cell intermingling are established at late stages of segmentation², perhaps because the EphA4 gene is a downstream transcriptional target of Krox-20 (ref. 14). Indeed, we found that three-somite embryos contain a scattered distribution of cells expressing truncated EphA4 or ephrin-B2 in the

hindbrain (Fig. 3b), whereas sorting to rhombomere boundaries has occurred in 10-somite embryos. To follow cell sorting, we co-injected RNA encoding truncated EphA4 and fluorescein dextran into one-cell embryos, and subsequently transplanted cells into the presumptive hindbrain¹⁵ of uninjected host embryos, which were later visualized by confocal microscopy. Interestingly, cell sorting occurred by initial movements along the antero-posterior axis of labelled neighbours to form a tight cluster of cells that then spread to form an aligned group (Fig. 5).

Classical studies in which cell populations are mixed with each other *in vitro* have shown that cells from different tissues sort out¹⁶, and that differential cell adhesion can lead to such sorting^{17–21}. This sorting is due to the preferential association of cells with similar adhesive properties during intermingling, and suggests that tissue-restricted cell adhesion molecules maintain organized patterns during development. Our findings show that activation of Eph receptors or ephrins can also drive cell sorting. In axonal growth cones, Eph-receptor activation leads to repulsion through collapse responses involving depolymerization of the actin cytoskeleton²². The stability of adhesive contacts between epithelial cells could be decreased by similar collapse responses, and/or by phosphorylation of cell adhesion molecules regulated by Eph receptors²³. An adhesive system that is uniformly expressed could thus be locally regulated by activation of Eph receptors or ephrins, perhaps leading to the larger intercellular spaces observed at rhombomere boundaries^{24,25}. Such responses will decrease the affinity of cells expressing ectopic ligand (truncated ephrin-B or truncated EphA4) for neighbours in which endogenous receptor (Eph receptor or ephrin, respectively) is activated. Cell sorting can thus be most simply explained as being due to differential cell affinity in which ligand-expressing cells

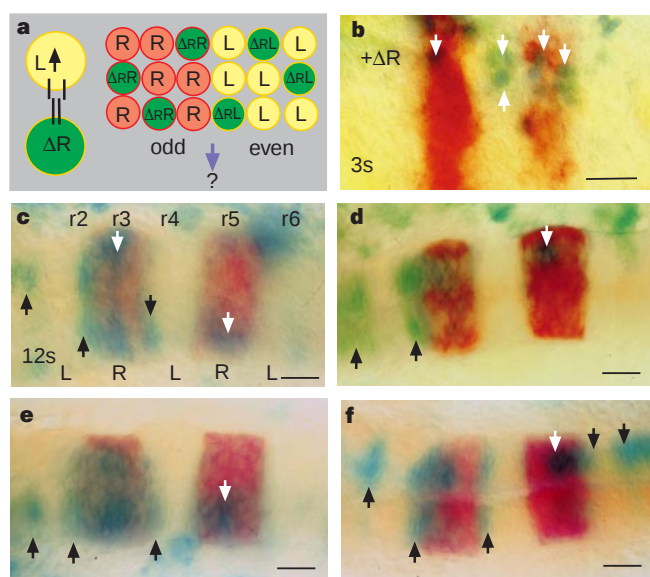


Figure 3 Cell sorting after mosaic expression of truncated EphA4. **a**, Left side: diagram illustrating activation of ephrin-B proteins (L) by truncated EphA4 (ΔR) which cannot itself transduce signals. Right side: following co-injection of *lacZ* and truncated *EphA4* RNA into one cell at the eight-cell stage, there will initially be a random distribution of cells co-expressing β -galactosidase and truncated receptor (green) in presumptive r3/r5 (orange) and r2/r4/r6 (yellow) territory. This will lead to mosaic activation of ephrin-B proteins in r2/r4/r6. **b–f**, Co-injection of *lacZ* and truncated *EphA4* RNA. **b**, Three-somite stage: expressing cells have a scattered distribution in the hindbrain (white arrows indicate examples). **c–f**, 15-somite stage: expressing cells have a scattered distribution in r3/r5, but sort to the boundaries of r2/r4/r6 (black arrows). Scale bars, 50 μ m.

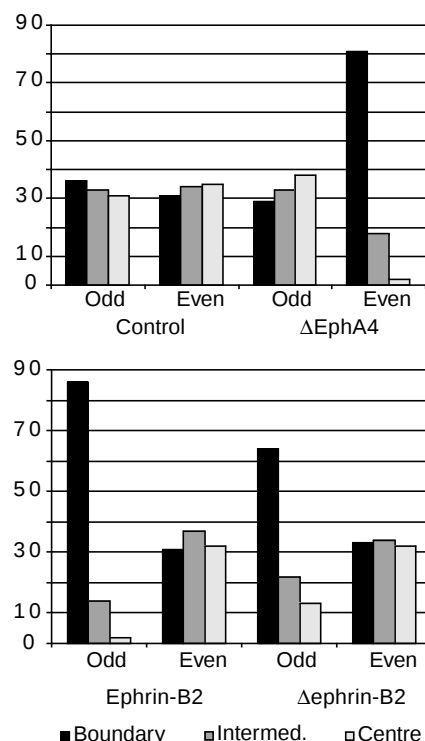


Figure 4 Sorting of cells adjacent to rhombomere boundaries. Quantification was carried out by measuring the area occupied by labelled cells in each of six zones along the antero-posterior axis of each rhombomere. This was used to calculate the proportion of cells in zones 1 plus 6 (adjacent to the boundary), 2 plus 5 (intermediate location) and 3 plus 4 (centre of the rhombomere) in odd- versus even-numbered rhombomeres. The bars indicate the average percentage of labelled cells at these locations in embryos expressing *lacZ* (control), $n = 10$; truncated EphA4, $n = 30$; ephrin-B2, $n = 20$; truncated ephrin-B2, $n = 10$.

minimize their contact with receptor-expressing cells. It is striking that the cells expressing ligand sort to boundaries, rather than forming aggregates within receptor-expressing territory. One possible explanation is that interaction of endogenous Eph receptors

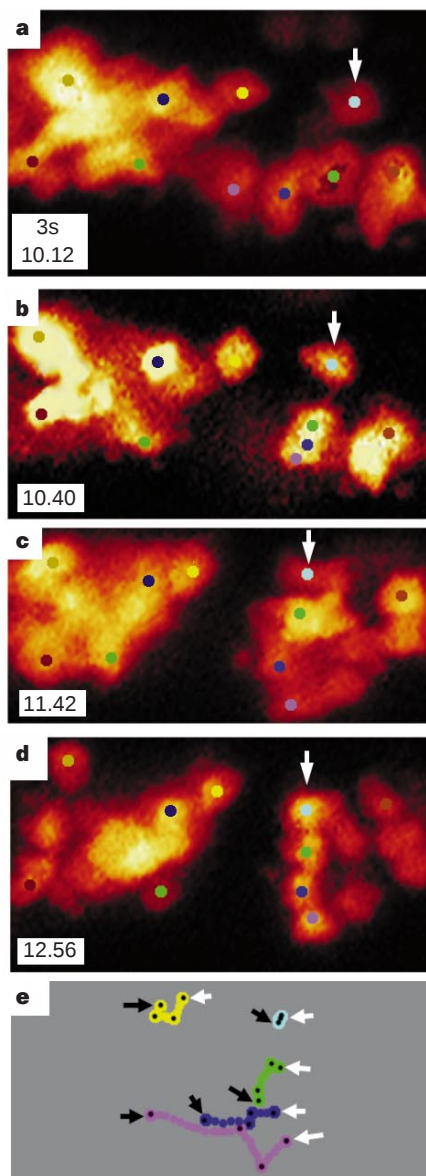


Figure 5 Time-lapse analysis of cell sorting. Cells from embryos co-injected at the one-cell stage with RNA encoding truncated EphA4 and fluorescein dextran were transplanted into the presumptive hindbrain of uninjected embryos. The resulting embryos were mounted at the two-somite stage and, at intervals, serial optical sections through the hindbrain were visualized using a confocal microscope. Most embryos had too few or too many expressing cells in the hindbrain, or failed to maintain an appropriate orientation throughout the required period, but in 1 out of 24 embryos the cell sorting process was visualized successfully. **a–d**, The coloured dots indicate specific cells that could be followed during the time-lapse analysis. Three cells expressing truncated EphA4 (green, dark blue and light blue dots) that are in contact with one another (**a**), cluster (**b**) and then spread to form an aligned group (**c, d**) indicated by the arrow. These movements generate a domain lacking labelled cells that at later stages was immediately anterior to the forming otic vesicle, so is likely to correspond to r4. The somite stage and/or time is indicated at the bottom left of each image. **e**, Diagram illustrating the movements of selected cells. Each coloured line indicates the movement of a cell (same colour as in **a–d**) from the start (black arrow) to the end (white arrow) of the time-lapse analysis. The black dots correspond to the location at each time point in **a–d**. The cell at the top right (light blue) was used as a fixed reference point.

and ephrins at rhombomere boundaries generates a region of cell–cell affinity similar to that of cells expressing exogenous ligand for neighbours expressing endogenous receptor. The finding that activation of Eph receptors or ephrins each lead to a cell-sorting response suggests that bi-directional activation at rhombomere boundaries may restrict cell intermingling. Support for this comes in experiments showing that expression of EphA4 and ephrin-B2 in adjacent groups of cells is sufficient to restrict intermingling (G.M., Q.X. and D.G.W., unpublished observations). The spatially restricted expression of Eph receptors and ephrins, and of cell-adhesion molecules, in many tissues during early development may therefore provide parallel, or perhaps synergistic, mechanisms for stabilizing patterns of cellular organization. □

Methods

Microinjection of RNA into zebrafish embryos. Preparation of RNA and microinjection were carried out as described elsewhere⁵. 0.1–0.5 nanolitre of *lacZ* RNA (0.2 mg ml⁻¹) plus RNA encoding Eph receptor or ephrin (0.2 mg ml⁻¹) was microinjected into one cell at the eight-cell stage. Detection using ephrin-Fc or Eph-Fc reagents¹¹ revealed that expression persists at similar levels to that of the endogenous proteins for at least 18 h. Constructs were subcloned into the SP64TK vector⁵, and encode the following sequences: full-length ephrin-B2, amino-acid residues 1–334; truncated ephrin-B2, residues 1–253; truncated EphA4, as described⁵; truncated EphB2 protein, residues 1–602.

Detection of β -galactosidase and *in situ* hybridization. Embryos were collected at the 15-somite stage, fixed in 4% paraformaldehyde, 0.25% glutaraldehyde in PBS for 15 min, then dechorionated. After washing three times for 10 min in 0.1% Triton X-100 in PBS, β -galactosidase was detected by staining in X-gal. The embryos were refixed in 4% paraformaldehyde in PBS for 15 min, then *in situ* hybridization was carried out²⁶ using Fast Red as substrate.

Transplantation and time-lapse microscopy. Embryos were injected at the one-cell stage with fluorescein dextran and RNA encoding truncated EphA4, and groups of cells were transplanted into the presumptive hindbrain¹⁵ of 50% epiboly-stage uninjected embryos. The resulting embryos were mounted under a coverslip in 3% methyl cellulose in 30% Danieau solution at the two-somite stage. At intervals, the fluorescein signal was visualized in serial optical sections through the hindbrain using a Leica confocal microscope. To display all labelled cells that are in different focal planes, serial sections through the hindbrain were merged using NIH Image software.

Analysis of Eph receptor and ephrin phosphorylation. Zebrafish embryos were injected at the one-cell stage with RNA encoding EphA4 or ephrin-B2, or truncated versions of these. At the 1,000-cell stage, embryos were dechorionated and dissociated into a single-cell suspension by incubation for 3 min in Ca²⁺/Mg²⁺-free Danieau solution including 5 mM EDTA. Cells were washed with PBS, then with L15 medium, and resuspended in L15 at 10⁵ cells per 30 μ l. Receptor- and ephrin-expressing cells were aggregated by mixing 30 μ l of each cell suspension and gently shaking for 10 min. The cells were then lysed in 300 μ l 1% Triton X-100, 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM sodium orthovanadate, 10 μ g ml⁻¹ aprotinin, 2 μ g ml⁻¹ leupeptin. To analyse EphA4 phosphorylation, EphA4 was immunoprecipitated²⁷, subjected to SDS–PAGE and western blotting, then probed with 4G10 anti-phosphotyrosine antibody, which was detected using ECL reagents (Amersham). The membrane was then stripped in 100 mM 2-mercaptoethanol, 2% SDS, 62.5 mM Tris-HCl pH 6.8 at 55 °C for 30 min, reprobed with anti-EphA4 antibody, and visualized as described above.

To analyse ephrin-B2 phosphorylation, ephrin-B2 was immunoprecipitated with 2 μ l anti-ephrin-B2 antibody (Santa Cruz) and subjected to SDS–PAGE. Western blotting and detection of phosphotyrosine were carried out as described above. After stripping of the membrane, ephrin-B2 protein was detected by sequential incubation with anti-ephrin-B2 antibody, then with biotinylated-protein G and finally with horseradish peroxidase conjugated to streptavidin. The ECL system was used to visualize the immunoreactive bands.

Received 3 December 1998; accepted 22 March 1999.

- Lawrence, P. A. The cellular basis of segmentation in insects. *Cell* **26**, 3–10 (1981).
- Fraser, S., Keynes, R. & Lumsden, A. Segmentation in the chick embryo hindbrain is defined by cell lineage restrictions. *Nature* **344**, 431–435 (1990).
- Guthrie, S., Prince, V. & Lumsden, A. Selective dispersal of avian rhombomere cells in orthotopic and heterotopic grafts. *Development* **118**, 527–538 (1993).
- Wizenmann, A. & Lumsden, A. Segregation of rhombomeres by differential chemoaffinity. *Mol. Cell.*

- Neurosci.* **9**, 448–459 (1997).
5. Xu, Q., Allard, G., Hilder, N. & Wilkinson, D. G. Expression of truncated Sek-1 receptor tyrosine kinase disrupts the segmental restriction of gene expression in the *Xenopus* and zebrafish hindbrain. *Development* **121**, 4005–4016 (1995).
 6. Xu, Q. & Wilkinson, D. G. Eph-related receptors and their ligands: mediators of contact dependent cell interactions. *J. Mol. Med.* **75**, 576–586 (1997).
 7. Flanagan, J. G. & Vanderhaeghe, P. The ephrins and Eph receptors in neural development. *Annu. Rev. Neurobiol.* **21**, 309–345 (1998).
 8. O'Leary, D. D. M. & Wilkinson, D. G. Eph receptors and ephrins in neural development. *Curr. Opin. Neurobiol.* **9**, 65–73 (1999).
 9. Holland, S. J. *et al.* Bidirectional signalling through the Eph-family receptor Nuk and its transmembrane ligands. *Nature* **383**, 722–725 (1996).
 10. Bruckner, K., Pasquale, E. B. & Klein, R. Tyrosine phosphorylation of transmembrane ligands for Eph receptors. *Science* **275**, 1640–1643 (1997).
 11. Gale, N. W. *et al.* Eph receptors and ligands comprise two major specificity subclasses, and are reciprocally compartmentalized during embryogenesis. *Neuron* **17**, 9–19 (1996).
 12. Eph Nomenclature Committee. Unified nomenclature for Eph family receptors and their ligands, the ephrins. *Cell* **90**, 403–404 (1997).
 13. Ellis, C. *et al.* A juxtamembrane autophosphorylation site in the Eph family receptor tyrosine kinase, Sek, mediates high affinity interaction with p59fyn. *Oncogene* **12**, 1727–1736 (1996).
 14. Theill, T. *et al.* Segmental expression of the EphA4 (Sek-1) gene is under direct transcriptional control of Krox-20. *Development* **125**, 443–452 (1998).
 15. Woo, K. & Fraser, S. E. Order and coherence in the fate map of the zebrafish nervous system. *Development* **121**, 2595–2609 (1995).
 16. Townes, P. L. & Holfreter, J. Directed movements and selective adhesion of embryonic amphibian cells. *J. Exp. Zool.* **128**, 53–120 (1955).
 17. Steinberg, M. S. Does differential adhesion govern self-assembly processes in histogenesis? Equilibrium processes and the emergence of a hierarchy among populations of embryonic cells. *J. Exp. Zool.* **173**, 395–434 (1970).
 18. Nose, A., Nagafuchi, A. & Takeichi, M. Expressed recombinant cadherins mediate cell sorting in model systems. *Cell* **54**, 993–1001 (1988).
 19. Friedlander, D. R., Mege, R. M., Cunningham, B. A. & Edelman, G. M. Cell sorting-out is modulated by both the specificity and amount of different cell adhesion molecules. *Proc. Natl Acad. Sci. USA* **86**, 7043–7047 (1989).
 20. Godt, D. & Tepass, U. *Drosophila* oocyte localization is mediated by differential cadherin-based adhesion. *Nature* **395**, 387–391 (1998).
 21. Gonzalez-Reyes, A. & St Johnston, D. The *Drosophila* A-P axis is polarized by the cadherin-mediated positioning of the oocyte. *Development* **125**, 3635–3644 (1998).
 22. Meima, L. *et al.* AL-1-induced growth cone collapse of rat cortical neurons is correlated with REK7 expression and rearrangement of the actin cytoskeleton. *Eur. J. Neurosci.* **9**, 177–188 (1997).
 23. Zisch, A. H. *et al.* Tyrosine phosphorylation of L1 family adhesion molecules: implication of the Eph kinase Cdk5. *J. Neurosci. Res.* **47**, 655–665 (1997).
 24. Lumsden, A. & Keynes, R. Segmental patterns of neuronal development in the chick hindbrain. *Nature* **337**, 424–428 (1989).
 25. Heyman, I., Kent, A. & Lumsden, A. Cellular morphology and extracellular space at rhombomere boundaries in the chick embryo hindbrain. *Dev. Dynamics* **198**, 241–253 (1993).
 26. Xu, Q., Hilder, N., Patient, R. & Wilson, S. W. Spatially regulated expression of three receptor tyrosine kinase genes during gastrulation in the zebrafish. *Development* **120**, 287–299 (1994).
 27. Irving, C., Nieto, M. A., DasGupta, R., Charnay, P. & Wilkinson, D. G. Progressive spatial restriction of Sek-1 and Krox-20 gene expression during hindbrain segmentation. *Dev. Biol.* **173**, 26–38 (1996).

Acknowledgements. We thank N. Gale and G. Yancopoulos for ephrin clones, M. Henkemeyer for the EphB2 clone and S. Fraser, J. P. Vincent, R. Krumlauf and P. Trainor for discussions. This work was supported by the MRC, an EC Biotechnology grant and an EMBO Fellowship to G.M.

Correspondence and requests for materials should be addressed to D.G.W. (e-mail: d-wilkin@nimr.mrc.ac.uk)

The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis

Patrick H. Maxwell^{*}, Michael S. Wiesener^{*},
Gin-Wen Chang[†], Steven C. Clifford[†], Emma C. Vaux[‡],
Matthew E. Cockman[‡], Charles C. Wykoff[‡],
Christopher W. Pugh[‡], Eamonn R. Maher[†]
& Peter J. Ratcliffe^{*‡}

^{*} Wellcome Trust Centre for Human Genetics, Windmill Road, Oxford OX3 7BN, UK
[†] Section of Medical and Molecular Genetics, Department of Paediatrics and Child Health, University of Birmingham, Birmingham B15 2TT, UK
[‡] Institute of Molecular Medicine, John Radcliffe Hospital, Oxford OX3 9DS, UK

Hypoxia-inducible factor-1 (HIF-1) has a key role in cellular responses to hypoxia, including the regulation of genes involved in energy metabolism, angiogenesis and apoptosis^{1–4}. The α subunits of HIF are rapidly degraded by the proteasome under normal conditions, but are stabilized by hypoxia⁵. Cobaltous ions or iron chelators mimic hypoxia, indicating that the stimuli may

interact through effects on a ferroprotein oxygen sensor^{6,7}. Here we demonstrate a critical role for the von Hippel-Lindau (VHL) tumour suppressor gene product pVHL in HIF-1 regulation. In VHL-defective cells, HIF α -subunits are constitutively stabilized and HIF-1 is activated. Re-expression of pVHL restored oxygen-dependent instability. pVHL and HIF α -subunits co-immunoprecipitate, and pVHL is present in the hypoxic HIF-1 DNA-binding complex. In cells exposed to iron chelation or cobaltous ions, HIF-1 is dissociated from pVHL. These findings indicate that the interaction between HIF-1 and pVHL is iron dependent, and that it is necessary for the oxygen-dependent degradation of HIF α -subunits. Thus, constitutive HIF-1 activation may underlie the angiogenic phenotype of VHL-associated tumours. The pVHL/HIF-1 interaction provides a new focus for understanding cellular oxygen sensing.

Enhanced glucose metabolism and angiogenesis are classical features of cancer^{8,9}, involving upregulation of genes that are normally induced by hypoxia. In addition to stimulation by the hypoxic microenvironment¹⁰, genetic alterations contribute to these effects^{8,9}. A striking example is von Hippel-Lindau (VHL) disease, a hereditary human cancer syndrome predisposing sufferers to highly angiogenic tumours¹¹. Constitutive upregulation of hypoxically inducible messenger RNAs encoding vascular endothelial growth factor (VEGF) and glucose transporter 1 (GLUT-1) in these tumour cells is correctable by re-expression of pVHL. A post-transcriptional mechanism has been proposed^{12,13}. We studied the involvement of pVHL in oxygen-regulated gene expression using ribonuclease protection analysis of two VHL-deficient renal carcinoma lines, RCC4 and 786-O. Eleven genes encoding products involved in glucose transport, glycolysis, high-energy phosphate metabolism and angiogenesis were examined; nine are induced by hypoxia in other mammalian cells and two (LDH-B and PFK-M) are repressed by hypoxia. None of these responses was seen in the VHL-defective cell lines. Responses to hypoxia were restored by stable transfection of a wild-type VHL gene, with effects ranging from a rather modest action of hypoxia (PFK-L and LDH-B) to substantial regulation (Fig. 1 shows results for RCC4 cells; similar effects were seen in 786-O cells, data not shown). These results indicate that the previously described upregulation of hypoxia-inducible mRNAs in VHL-defective cells^{12,13} extends to a broad range of oxygen-regulated genes and involves a constitutive 'hypoxia pattern' for both positively and negatively regulated genes.

As a number of these genes (VEGF, GLUT-1, AK-3, ALD-A, PGK-1, PFK-L and LDH-A) contain hypoxia-response elements (HREs) which bind HIF-1 and/or show altered expression in cells lacking HIF-1 (refs 2, 14 and references therein), this survey of expression in VHL-defective cells prompted us to look for effects of pVHL on HIF-1 and HRE function. RCC4 cells were co-transfected with reporter plasmids which did or did not contain HREs from the mouse phosphoglycerate-kinase-1 or erythropoietin (Epo) enhancers, and either the VHL-expression plasmid pcDNA3-VHL or an empty vector. pVHL suppressed HRE activity in normoxic cells and restored induction by hypoxia (Fig. 2a). Similar results were obtained by sequential stable transfection of RCC4 cells with an HRE reporter followed by pcDNA3-VHL (data not shown). HIF-1 itself was examined by electrophoretic mobility shift assay (EMSA), which showed a constitutive HIF-1 DNA-binding species in VHL-deficient RCC4 cells, with restoration of the normal hypoxia-inducible pattern in RCC4 cells stably transfected with pcDNA3-VHL (RCC4/VHL) (Fig. 2b). In other cells, HIF-1 activation by hypoxia involves a large increase in HIF-1 α abundance from low basal levels in normoxia^{1,15}. Western blotting of whole-cell extracts showed that RCC4 cells express constitutively high levels of both HIF-1 α and a related molecule, HIF-2 α (also termed EPAS-1, HRE, HLF and MOP2) which is normally regulated in a similar way¹⁶ (Fig. 2c). Constitutively high levels of these proteins were found in eight other VHL-defective cell lines, in contrast to the renal

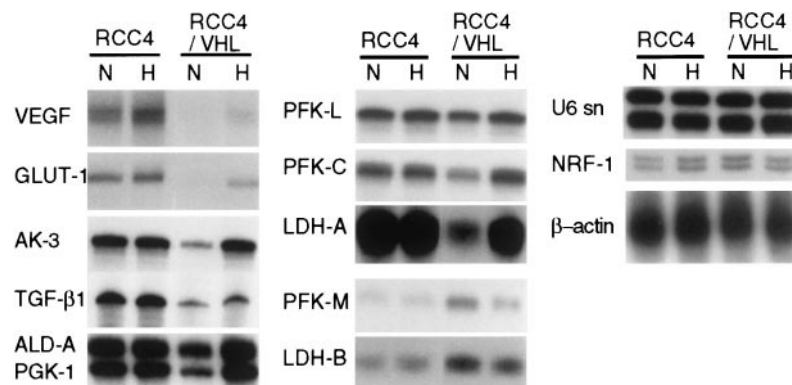


Figure 1 Effect of pVHL on oxygen-regulated gene expression. mRNA analysis of RCC4 cells and stable transfectant expressing pVHL (RCC4/VHL). N, normoxia; H, hypoxia (1% O₂, 16 h). VEGF, vascular endothelial growth factor; GLUT-1, glucose transporter 1; AK-3, adenylate kinase 3; TGF-β1, transforming growth factor-β1; ALD-A, aldolase A; PGK-1, phosphoglyceratekinase 1; PFK, phos-

phofructokinase; LDH, lactate dehydrogenase. U6 small nuclear (sn) RNA was used as an internal control. Also illustrated are two genes not influenced by VHL status or hypoxia; nuclear respiratory factor 1 (NRF-1) and β-actin. Amount of RNA analysed is detailed in Table S1 of the Supplementary Information.

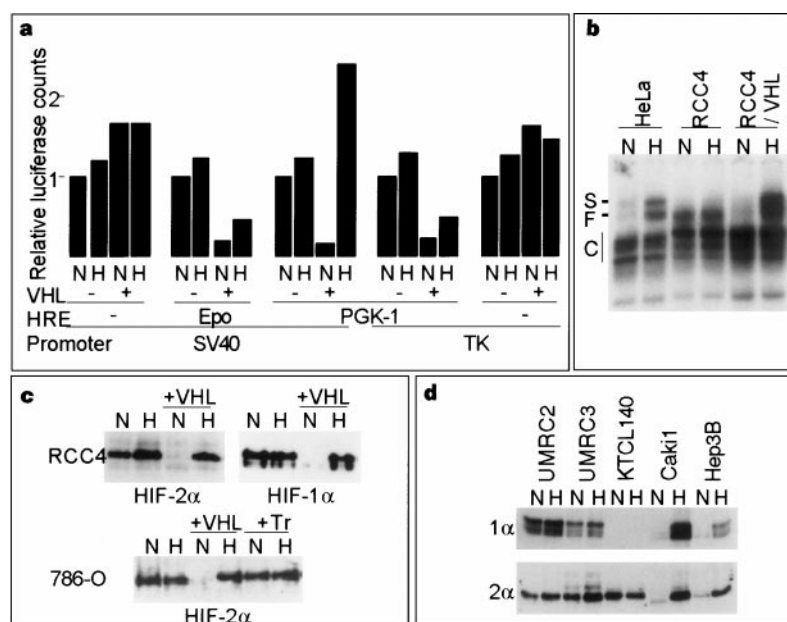


Figure 2 Effect of pVHL on HIF-1 and HRE activity. **a**, Representative transient transfections of RCC4 cells with VHL expression vector (+) or empty vector, and luciferase reporter genes containing no HRE, or HREs from the PGK-1 or erythropoietin (Epo) genes linked to SV40 or TK promoters. N, normoxia; H, hypoxia (0.1% O₂, 24 h—results were similar with 1% O₂). **b**, EMSA using the Epo HRE. N, normoxia; H, hypoxia (1% O₂, 4 h). In HeLa and RCC4/VHL, HIF-1 is a doublet (S, slower and F, faster mobility). RCC4 cells contain only faster mobility HIF-1, with equivalent levels in normoxia and hypoxia. Constitutive binding

species are indicated (C). **c**, Immunoblots of whole cell extracts for HIF α-subunits. Upper panels; RCC4 and RCC4/VHL cells (+VHL). Lower panel; 786-O cells stably transfected with vector alone, a full length VHL gene (+VHL), or a truncated VHL gene (1-115; +Tr). HIF-1α was not detected in 786-O cells. **d**, Immunoblot of UMRC2, UMRC3 and KTCL140 (renal carcinoma lines with VHL mutations³⁰), Caki-1 (renal carcinoma line expressing wild-type pVHL) and the hepatoma line Hep3B.

carcinoma line Caki-1 (which expresses pVHL normally¹⁷) and a wide range of previously reported cell lines (Fig. 2d and Supplementary Information S2). Certain VHL-defective cells (for example, 786-O, KTCL140) expressed HIF-2α at a high constitutive level but did not express detectable HIF-1α (Fig. 2c, d and Supplementary Information S2). Examination of stable transfectants of RCC4 and 786-O cells demonstrated that expression of the wild-type, but not truncated, VHL gene restored regulation of HIF α-subunits by oxygen without affecting the levels of mRNA encoding either subunit (data not shown).

To investigate the role of pVHL in HIF-1 regulation we tested for interactions between HIF α-subunits and pVHL using a combination of hypoxia and/or proteasomal blockade to induce HIF α-subunits. Anti-pVHL immunoprecipitates of extracts from protea-

somally blocked RCC4/VHL cells, but not RCC4 cells, contained both HIF-1α and HIF-2α (Fig. 3a). Similar results were obtained with hypoxia in the absence of proteasomal blockade. In the inverse reactions, immunoprecipitating antibodies to HIF-2α or HIF-1α co-precipitated pVHL, although a smaller proportion of the total was captured (Fig. 3b). Anti-pVHL immunoprecipitations also demonstrated the interaction in HeLa cells, which express pVHL normally (Fig. 3c). We next tested whether pVHL is incorporated in the HIF-1 DNA-binding complex. Addition of antibodies raised against pVHL (but not control antibodies) to nuclear extract from RCC4/VHL cells and HeLa cells produced a change in mobility, which was not observed with VHL-defective RCC4 cells (Fig. 3d). HIF-1 migrated as two species. Only the slower-mobility HIF-1 species was shifted by anti-pVHL, whereas both species were shifted

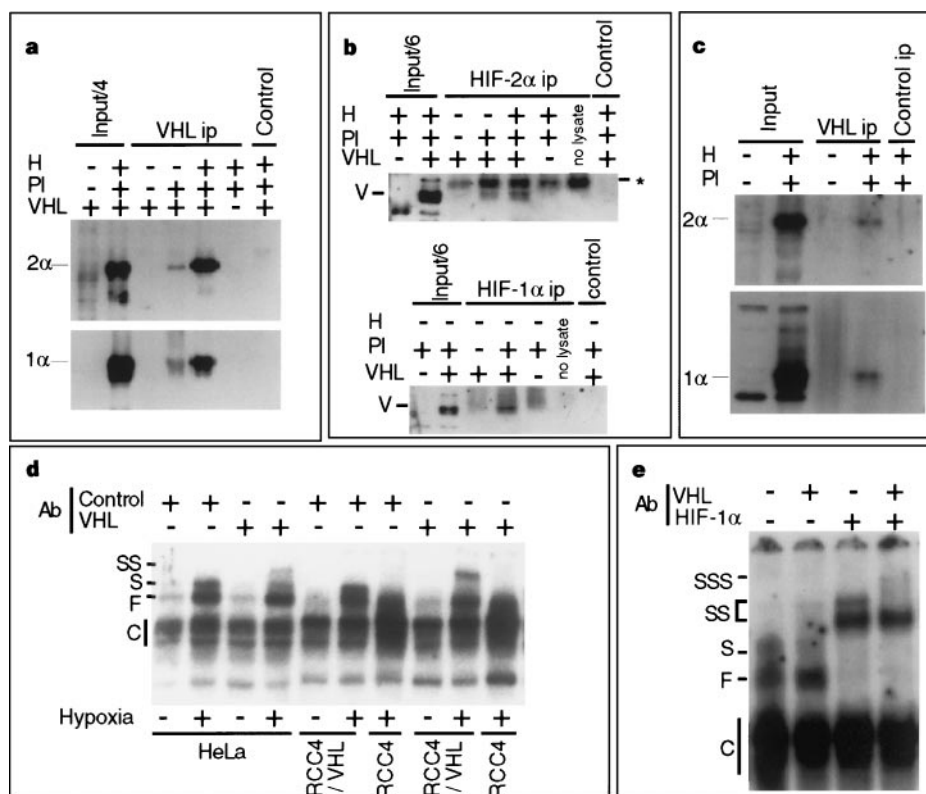


Figure 3 Association of pVHL with HIF-1. **a**, Immunoblots for HIF α -subunits (2 α , 1 α) of IG32 (VHL ip) and control immunoprecipitates (using VG-7be) of RCC4/VHL (VHL+) and RCC4 (VHL-) cells exposed (4 h) to normoxia or hypoxia (1% O₂; H+) with or without proteasomal inhibition (PI+). Aliquots of selected input lysates were also immunoblotted. **b**, Immunoprecipitation of RCC4/VHL (VHL+) and RCC4 (VHL-) extracts with polyclonal antibodies to HIF α -subunits or normal rabbit immunoglobulin (control ip) followed by immunoblotting for pVHL (V). A cross-reacting species arose from the HIF-2 α antibody (asterisk). **c**, Immunoprecipitation of HeLa extracts with IG32 (VHL_{ip}) or pAb419 (control) followed by

immunoblotting for HIF α -subunits. **d**, Anti-pVHL supershifts. IG32 (VHL Ab+) or VG-7be (Control Ab+) was added to binding reactions of nuclear extracts from normoxic or hypoxic (1% O₂, 4 h; H+) cells. Anti-pVHL supershifted (SS) the slower HIF-1 species (S) in HeLa and hypoxic RCC4/VHL cells. No supershift was seen with RCC4 cells, which lack pVHL. **e**, Anti-pVHL, anti-HIF-1 α and combination supershifts in HeLa cells. Anti-pVHL (IG32, +VHL Ab) supershifted slower mobility HIF-1. Anti-HIF-1 α (clone 54) supershifted both components (SS). Addition of both antibodies 'super-super-shifted' (SSS) slower mobility HIF-1.

by anti-HIF-1 α (Fig. 3e). Similar results were obtained in other cell lines (Hep3B, Caki-1, MRC5-V2 and 293; data not shown). Furthermore, whereas RCC4/VHL, HeLa and other cells contained both HIF-1 species, RCC4 extracts contained only the faster-mobility species (Figs 3d, Fig. 2b). Thus, VHL-defective cells lack the slower-mobility species which is restored by re-expression of pVHL, and shifted by anti-pVHL. This indicates that the DNA-binding HIF-1 doublet arises from two species—containing or not containing pVHL. Combination supershift analysis confirmed that the slower-mobility species contained both HIF-1 α and pVHL (Fig. 3e).

HIF-1 activation by hypoxia is mimicked by cobaltous ions and iron chelation^{6,7}. We therefore tested whether the pVHL/HIF-1 interaction was regulated by these stimuli. Proteasomal blockade induces an HIF-1 DNA-binding complex in normoxic cells¹⁸; comparison of this normoxic complex with EMSA of hypoxic cells with or without proteasomal inhibitors showed a similar shift and anti-pVHL supershift (Fig. 4a and data not shown). Together with immunoprecipitation data, this indicates that the interaction with pVHL occurs in both normoxic and hypoxic cells. In contrast, EMSA analysis of RCC4/VHL cells treated with cobalt and the iron chelator desferrioxamine (DFO) demonstrated only the faster mobility HIF-1. This did not supershift with anti-pVHL, indicating that the pVHL/HIF-1 complex could not form in cells exposed to these stimuli. Similar results were obtained in other cell types and are consistent with hitherto unexplained mobility differences between previous analyses of HIF-1 from cobalt- or DFO-versus hypoxia-stimulated cells⁷, indicating that this is a general

effect. Treatment with DFO 4 h before hypoxia prevented the formation of the pVHL/HIF-1 complex (Fig. 4b). Addition of iron chelators could not break the pVHL/HIF-1 complex *in vitro*, whereas addition of *in vitro*-translated wild-type pVHL (but not a truncated pVHL) could restore the slower-mobility species to nuclear extracts of proteasomally blocked, normoxic and hypoxic RCC4 cells, but not to cells treated with DFO or cobalt (Fig. 4c, and Fig. S1a of Supplementary Information). Immunoprecipitation studies also indicated that the interaction between HIF-1 and pVHL is iron-dependent. Whereas both HIF-1 α and HIF-2 α were contained in anti-pVHL immunoprecipitates from hypoxic RCC4/VHL cells, neither was contained in precipitates from DFO- or cobalt-treated cells (Fig. 4d). The iron-dependent interaction between HIF α -subunits and pVHL may be direct or indirect. However, *in vitro*-translated wild-type pVHL did not bind to an *in vitro*-translated HIF-1 DNA-binding complex (Fig. S1b in Supplementary Information), in contrast to the interaction with RCC4 extracts (Fig. 4c), indicating that an additional factor or modification of HIF-1 not represented in rabbit reticulocyte lysates is necessary for the association.

Normally, HIF α -subunits are targeted for rapid degradation in normoxic cells by a proteasomal mechanism operating on an internal oxygen-dependent-degradation (ODD) domain⁵. Our data suggest that pVHL might be required for this process—a possibility which would be consistent with recent data that pVHL forms a multiprotein complex (containing Cul-2 and elongins B and C) which has homology with ubiquitin-ligase/proteasome-targeting complexes in yeast^{19,20}. When cells were switched from

hypoxia to normoxia with addition of cycloheximide, HIF α -subunits decayed with a half-life of about 5 min in wild-type VHL transfectants, compared with ~ 60 min in the VHL-defective RCC4 and 786-O cells, thus confirming a strong effect of pVHL on stability (Fig. 5a). Moreover, functional studies of Gal4 chimaeras containing the HIF-1 α ODD domain demonstrated a striking dependence of the isolated ODD domain on pVHL (Fig. 5b).

These experiments define a function for pVHL in the regulation of HIF-1. Given recent demonstrations of the importance of HIF-1 in tumour angiogenesis^{3,21}, constitutive HIF-1 activation is clearly consistent with the angiogenic phenotype of VHL disease. Whether it is a sufficient explanation for oncogenesis is less clear. HIF-1 mediates gene activation not only by hypoxia, but also by growth factors such as insulin and insulin-like growth factor-1 (ref. 22). HIF-1 targets such as molecules involved in enhanced glucose metabolism and angiogenesis^{8,9,14} are classically upregulated (by different mechanisms) in many forms of cancer, supporting an

important role in tumour progression, although their role in the initiation of oncogenesis is less clear. pVHL is probably a multi-functional protein which could have other tumour suppressor actions^{11,23,24}. One possibility is that other gene products could be targeted in a similar manner to HIF α -subunits; however, comparison of anti-pVHL immunoprecipitates from metabolically labelled RCC4/VHL cells with and without proteasomal inhibitors has so far demonstrated only two species: HIF-1 α and HIF-2 α (G.-W.C., unpublished observations).

As both pVHL and HIF-1 are widely expressed, it is likely that the physiological role of pVHL in HIF-1 regulation is general. It is not yet clear whether pVHL has actions on oxygen-regulated gene expression other than through HIF-1. Stabilization of hypoxia-inducible mRNAs has been reported in VHL-deficient cells^{12,13}. This might represent an independent action of pVHL. However, regulation of these RNAs is commonly abolished in HIF-1-deficient cells^{3,25}, so the mRNA stability factors could lie downstream of

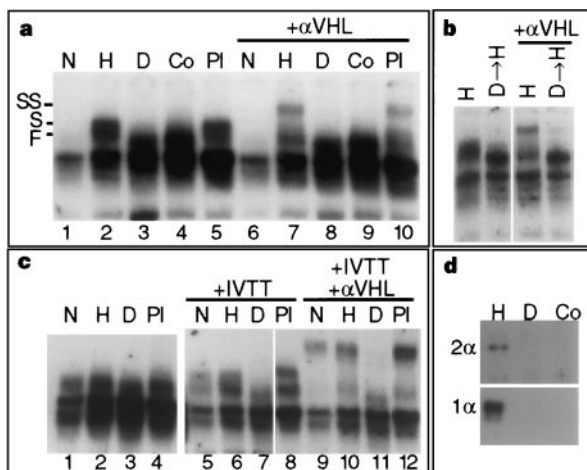


Figure 4 Effect of cobaltous ions and iron chelation on the pVHL/HIF-1 interaction. **a**, EMSA and supershift analysis of RCC4/VHL cells exposed (4 h) to normoxia (N), hypoxia (H; 1% O₂), DFO (100 μ M), cobaltous chloride (Co, 100 μ M) or proteasomal inhibition (PI). In lanes 6–10, IG32 was added (+ α VHL). **b**, EMSA and supershift analysis of RCC4/VHL cells subjected to hypoxia (H; 1% O₂ for 8 h) or DFO (100 μ M) with hypoxia (D $\rightarrow$ H; cells exposed to DFO for 8 h, 4 h

normoxia, then 4 h hypoxia). **c**, EMSA of RCC4 cells showing only faster mobility HIF-1 (lanes 1–4). Addition of *in vitro* transcribed/translated pVHL alone (lanes 5–8, +IVTT), or together with IG32 (lanes 9–12, +IVTT+ α VHL); slower mobility HIF-1 forms and supershifts in cells exposed to normoxia, hypoxia and PI, but not DFO. **d**, HIF α immunoblots of anti-pVHL immunoprecipitates of RCC4/VHL cells.

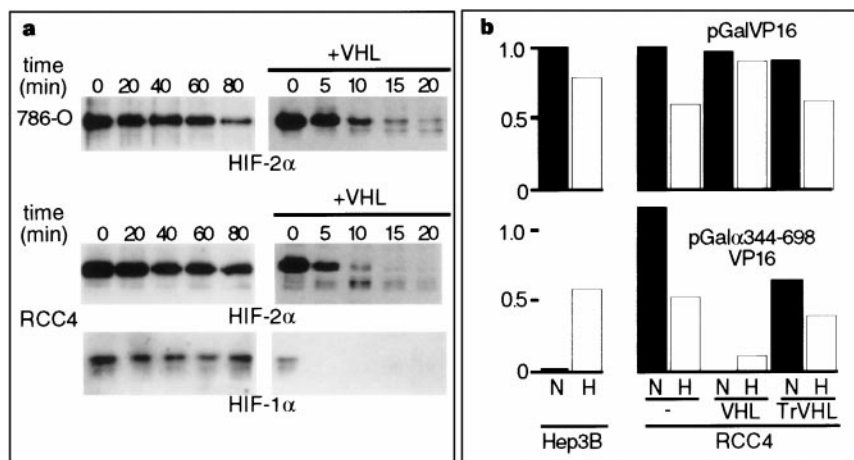


Figure 5 Effect of pVHL on HIF α stability and ODD domain function. **a**, Western analysis of HIF α -subunit stability in cells lacking VHL (RCC4, 786-O) and stable transfectants expressing pVHL (+VHL). Cells were incubated in hypoxia (4 h), then moved to normoxia (time 0) with addition of cycloheximide (100 μ M) and harvested up to 80 min later. **b**, Representative functional assay of the HIF-1 α ODD domain. Hep3B or RCC4 cells were transfected with Gal4 reporter pUAS-tk-Luc, and either pGalVP16 (upper panel) encoding the Gal4 DNA-binding domain fused to the VP16 activation domain, or pGal α 344-698VP16 (lower panel) which includes

HIF-1 α amino acids 344–698 (containing the entire ODD domain⁶). RCC4 cells were co-transfected with pcDNA3 (–), pcDNA3-VHL (VHL) or pcDNA3-VHL.103FS (TrVHL). After transfection, cells were divided for 24 h incubation in normoxia (N) or hypoxia (H; 0.1% O₂). Corrected luciferase counts are shown, normalized to the normoxic value with pGalVP16 or pGalVP16+pcDNA3. The HIF-1 α domain confers suppression and hypoxic regulation in Hep3B cells but not RCC4 cells, where re-expression of pVHL restores these properties.

HIF-1. Positive and negative effects on HIF-1 activation by iron chelators, hydrogen peroxide and redox active agents have indicated that the underlying oxygen sensing mechanism may involve oxygen-dependent generation of partially reduced oxygen radical species by redox active iron centre(s)^{2,6,7,15,26}. Such system(s) could act distantly on a transduction pathway or might directly modify the ODD domains of HIF α through a caged radical system, as proposed for iron sensing in the regulation of IRP2 (ref. 27). Cobaltous ions activate HIF-1 in inverse relation to iron availability, indicating that there may be competition for incorporation into a metal centre²⁸.

Our data support a model in which VHL/HIF complexes form in normoxic cells and target HIF α subunits for destruction. In hypoxia, degradation is suppressed despite complex formation, perhaps because a critical targeting modification of the HIF α ODD domain cannot occur without oxygen. Cobaltous ions and desferrioxamine prevent formation of the VHL complex, providing a different mechanism for stabilization of HIF-1 and potentially explaining why activation by these stimuli is relatively resistant to oxygen and certain other radical-generating processes^{6,16,29}. The iron dependence of complex formation could indicate that an iron-containing protein is an essential component of the complex, perhaps involved in local generation of an oxygen-sensing signal. □

Methods

Cells and transfections. 786-O cells expressing pVHL, truncated pVHL (amino acids 1–115), or empty vector¹⁷ were a gift from W. G. Kaelin. RCC4 cells were a gift from C. H. C. M. Buys. Other RCC lines were provided by M. Lerman. HeLa and Hep3B cells were from ECACC. RCC4/VHL was obtained by transfection with pcDNA3-VHL and G418 selection. Cells were plated in medium lacking G418 24 h before experiments, which were performed on 75 cm² dishes approaching confluence. Proteasomal inhibition was with 100 μ M calpain inhibitor I and 10 μ M N-carbobenzoxyl-L-leucyl-L-norvalinal. Transient transfections were by electroporation. Transfected cells were split for parallel normoxic and hypoxic incubation (Napco 7001, Precision Scientific). Luciferase reporter gene activity was corrected for transfection efficiency by assay of β -galactosidase expression from the co-transfected control plasmid pCMV- β Gal.

RNA analysis. Total RNA was extracted and analysed by ribonuclease protection. Riboprobe details are given in Table S1 of the Supplementary Information.

Plasmid constructions. pcDNA3-VHL contained nucleotides 214–855 of GenBank accession no. L15409 in pcDNA3 (Invitrogen). pcDNA3-VHL.103FS was made using site-directed mutagenesis to delete nucleotides 522–523. HRE reporter genes were based on pGL3-basic (Promega) or pPUR (Clontech) and contained either a minimal SV40 promoter or a minimal (–40 bp) thymidine kinase promoter linked to a firefly luciferase gene (details of HREs are in Supplementary Information S3). pGalVP16 encoded the Gal4 DNA-binding domain (amino acids 1–147) linked in-frame to the activation domain (amino acids 410–490) from herpes simplex virus protein 16; pGal α 344–698VP16 encoded the indicated amino acids of HIF-1 α between those domains. Plasmid pUAS-tk-Luc contained two copies of the Gal4 binding site linked to a thymidine-kinase-promoted luciferase reporter gene.

Cell lysis, immunoblotting and immunoprecipitation. Whole cell extracts were prepared by homogenization in denaturing conditions and aliquots immunoblotted for HIF α -subunits with 28b (anti-HIF-1 α) and 190b (anti-HIF-2 α) as described¹⁶, or using clone 54 (anti-HIF-1 α , Transduction Laboratories). For immunoprecipitation, lysis was performed in 100 mM NaCl, 0.5% Igepal CA630, 20 mM Tris-HCl (pH 7.6), 5 mM MgCl₂ and 1 mM sodium orthovanadate with aprotinin (10 μ g ml^{–1}), ‘Complete’ protease inhibitor (Boehringer) and 1.0 mM 4-(2-aminoethyl)benzene sulphonyl fluoride for 30 min on ice. After clearance by centrifugation, 120 μ g aliquots of lysate were incubated for 2 h at 4 °C with 4 μ g affinity-purified anti-HIF-2 α polyclonal antibodies (raised against a bacterially expressed fusion protein including amino acids 535–631) or 4 μ g ammonium sulphate precipitated anti-HIF-1 α polyclonal antibodies (raised against an immunogen including amino acids 530–652) in parallel with normal rabbit immunoglobulin (control), or alternatively with 0.7 μ g anti-pVHL antibody (IG32, Pharmingen) or control (antibody to SV40 T antigen, pAb419, a gift from E. Harlow or antibody to

VEGF, VG-7be, a gift from H. Turley). 10 μ l conjugated agarose beads pre-blocked with 20 mg ml^{–1} BSA was added and lysates incubated for 2 h with rocking. Pellets were washed five times, eluted with sample buffer, and divided into 2–6 aliquots for immunoblotting.

Electrophoretic mobility shift and supershift assays. We prepared nuclear extracts using a modified Dignam protocol and incubated 5 μ g (HeLa) or 7.5 μ g (RCC4) with a ³²P-labelled 24-bp oligonucleotide probe (sense strand; 5'-GCCCTACGTGCTGCCTCGCATGGC-3') from the mouse Epo 3' enhancer as described²⁵. For supershift assays, 0.5 μ g IG32, VG-7be (isotype and subclass matched control for IG32) or clone 54 (anti-HIF-1 α) was added and reactions were incubated for 4 h at 4 °C before electrophoresis. *In vitro* transcription translations of pcDNA3-VHL and pcDNA3-VHL.103FS were done using reticulocyte lysate (Promega); 1 μ l of a 1:5 dilution in PBS was added to binding reactions 2 h before electrophoresis or addition of antibody.

Received 11 March; accepted 13 April 1999.

- Wang, G. L., Jiang, B.-H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O₂ tension. *Proc. Natl Acad. Sci. USA* **92**, 5510–5514 (1995).
- Bunn, H. F. & Poyton, R. O. Oxygen sensing and molecular adaptation to hypoxia. *Physiol. Rev.* **76**, 839–885 (1996).
- Carmeliet, P. et al. Role of HIF-1 α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. *Nature* **394**, 485–490 (1998).
- An, W. G. et al. Stabilization of wild-type p53 by hypoxia-inducible factor 1 α . *Nature* **392**, 405–408 (1998).
- Huang, L. E., Gu, J., Schau, M. & Bunn, H. F. Regulation of hypoxia-inducible factor 1 α is mediated by an oxygen-dependent domain via the ubiquitin-proteasome pathway. *Proc. Natl Acad. Sci. USA* **95**, 7987–7992 (1998).
- Goldberg, M. A., Dunning, S. P. & Bunn, H. F. Regulation of the erythropoietin gene: evidence that the oxygen sensor is a heme protein. *Science* **242**, 1412–1415 (1988).
- Wang, G. L. & Semenza, G. L. Desferrioxamine induces erythropoietin gene expression and hypoxia-inducible factor 1 DNA-binding activity: implications for models of hypoxia signal transduction. *Blood* **82**, 3610–3615 (1993).
- Warburg, O. *The Metabolism of Tumours* (Arnold Constable, London, 1930).
- Hanahan, D. & Folkman, J. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell* **86**, 353–364 (1996).
- Shweiki, D., Itin, A., Soffer, D. & Keshet, E. Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. *Nature* **359**, 843–845 (1992).
- Kaelin, W. G. & Maher, E. R. The VHL tumour-suppressor gene paradigm. *Trends Genet.* **14**, 423–426 (1998).
- Gnarra, J. R. et al. Post-transcriptional regulation of vascular endothelial growth factor mRNA by the product of the VHL tumour suppressor gene. *Proc. Natl Acad. Sci. USA* **93**, 10589–10594 (1996).
- Iliopoulos, O., Levy, A. P., Jiang, C., Kaelin, W. G. Jr & Goldberg, M. A. Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. *Proc. Natl Acad. Sci. USA* **93**, 10595–10599 (1996).
- Dang, C. V. & Semenza, G. L. Oncogenic alterations of metabolism. *Trends Biochem. Sci.* **24**, 68–72 (1999).
- Huang, L. E., Arany, Z., Livingston, D. M. & Bunn, H. F. Activation of hypoxia-inducible transcription factor depends primarily on redox-sensitive stabilization of its α subunit. *J. Biol. Chem.* **271**, 32253–32259 (1996).
- Wiesener, M. S. et al. Induction of endothelial PAS domain protein-1 by hypoxia: characterization and comparison with hypoxia-inducible factor-1 α . *Blood* **92**, 2260–2268 (1998).
- Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W. G. Jr Tumour suppression by the human von Hippel-Lindau gene product. *Nature Med.* **1**, 822–826 (1995).
- Salceda, S. & Caro, J. Hypoxia-inducible factor 1 (HIF-1 α) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. *J. Biol. Chem.* **272**, 22642–22647 (1997).
- Pause, A. et al. The von Hippel-Lindau tumor-suppressor gene product forms a stable complex with human CUL-2, a member of the Cdc53 family of proteins. *Proc. Natl Acad. Sci. USA* **94**, 2156–2161 (1997).
- Loneragan, K. M. et al. Regulation of hypoxia-inducible mRNAs by the von Hippel-Lindau tumor suppressor protein requires binding to complexes containing elongins B/C and Cul2. *Mol. Cell Biol.* **18**, 732–741 (1998).
- Maxwell, P. H. et al. Hypoxia inducible factor-1 modulates gene expression in solid tumours and influences both angiogenesis and tumor growth. *Proc. Natl Acad. Sci. USA* **94**, 8104–8109 (1997).
- Zelzer, E. et al. Insulin induces transcription of target genes through the hypoxia-inducible factor HIF-1 α /ARNT. *EMBO J.* **17**, 5083–5094 (1998).
- Duan, D. R. et al. Inhibition of transcription elongation by the VHL tumor suppressor protein. *Science* **269**, 1402–1406 (1995).
- Ohh, M. et al. The von Hippel-Lindau tumor suppressor protein is required for proper assembly of an extracellular fibronectin matrix. *Mol. Cell* **1**, 959–968 (1998).
- Wood, S. M. et al. Selection and analysis of a mutant cell line defective in the hypoxia-inducible factor-1 α -subunit (HIF-1 α). *J. Biol. Chem.* **273**, 8360–8368 (1998).
- Fandrey, J., Frede, S. & Jelkmann, W. Role of hydrogen peroxide in hypoxia-induced erythropoietin production. *Biochem. J.* **303**, 507–510 (1994).
- Iwai, K. et al. Iron-dependent oxidation, ubiquitination, and degradation of iron regulatory protein 2: implications for degradation of oxidized proteins. *Proc. Natl Acad. Sci. USA* **95**, 4924–4928 (1998).
- Ho, V. T. & Bunn, H. F. Effects of transition metals on the expression of the erythropoietin gene: further evidence that the oxygen sensor is a heme protein. *Biochem. Biophys. Res. Commun.* **223**, 175–180 (1996).
- Chandel, N. S. et al. Mitochondrial reactive oxygen species trigger hypoxia-induced transcription. *Proc. Natl Acad. Sci. USA* **95**, 11715–11720 (1998).
- Gnarra, J. R. et al. Mutations of the VHL tumour suppressor gene in renal carcinoma. *Nature Genet.* **7**, 85–90 (1994).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank W. Kaelin, C. Buys and M. Lerman for cell lines, and N. Proudfoot, A. Harris, D. Gillespie, J. O'Rourke, Y.-M. Tian and L. Nicholls. Financial support was from the Wellcome Trust, the Barnes Trust, the Deutsche Forschungsgemeinschaft, the Cancer Research Campaign, Action Research and the Medical Research Council.

Correspondence and requests for materials should be addressed to P.J.R. (e-mail: peter.ratcliffe@imm.ox.ac.uk).

Mammalian Srb/Mediator complex is targeted by adenovirus E1 A protein

Thomas G. Boyer^{*†}, Michelle E. D. Martin^{*}, Emma Lees[‡], Robert P. Ricciardi[§] & Arnold J. Berk^{*}

^{*} Molecular Biology Institute, UCLA, Los Angeles, California 90095-1570, USA

[‡] DNAX, Palo Alto, California 94304-1104, USA

[§] Department of Microbiology, School of Dental Medicine, and Department of Biochemistry/Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Adenovirus E1A proteins prepare the host cell for viral replication, stimulating cell cycling and viral transcription through interactions with critical cellular regulatory proteins such as RB^{1,2} and CBP³. Here we show that the E1A zinc-finger domain that is required to activate transcription of viral early genes binds to a host-cell multiprotein complex containing homologues of yeast Srb/Mediator proteins^{4,5}. This occurs through a stable interaction with the human homologue of *Caenorhabditis elegans* SUR-2, a protein required for many developmental processes in the nematode⁶. This human Srb/Mediator complex stimulates transcription *in vitro* in response to both the E1A zinc-finger and the herpes simplex virus VP16 activation domains. Interaction with human Sur-2 is also required for transcription to be activated by the activation domain of a transcription factor of the ETS-family in response to activated mitogen-activated protein (MAP) kinase.

To identify potential targets of the E1A zinc-finger within conserved region 3 (CR3), protein fractions AB, CB and D (Fig. 1a), which collectively support CR3-activated transcription *in vitro*⁷, were passed over a glutathione S-transferase (GST)–E1A CR3 column; an equivalent column was prepared with a transcriptionally defective CR3 point mutant⁸. Under non-stringent conditions (0.1 M KCl), a complex profile of polypeptides from the AB and CB fractions bound, with no apparent differences between the wild-type and the mutant column. However, a ~150K protein (relative molecular mass (M_r) ~150,000) from the D-fraction bound specifically to the wild-type and not the mutant column. Under stringent conditions (0.7 M KCl), this protein bound to CR3 point mutants which are active for stimulating viral early transcription (mutants H158F and T178S), but not to mutants with inactive activation domains (H160Y, Y175F, M176K, C154S, C171S, L173F and R177K; Fig. 1b, c; data not shown for all mutants).

A sequenced peptide⁹ from the CR3-binding protein matched a human expressed sequence tag (EST), and antibodies raised against an 88-residue EST-encoded peptide specifically immunoreacted with the CR3-binding protein (Fig. 1c). RNA blot analysis using the EST as the probe revealed an ~4-kilobase (kb) poly(A)⁺ RNA expressed in each of eight mammalian tissues analysed (data not shown). A full-length complementary DNA was isolated that encoded a novel human polypeptide of 1,364 amino acids with significant homology to *C. elegans* SUR-2 (Fig. 1e), a ubiquitous protein of unknown function that acts late in a receptor-tyrosine kinase–Ras–Raf–MAP kinase signalling pathway required for several developmental processes, including vulval cell development⁶. To test the binding of this human SUR-2 homologue (hSur-2) to E1A CR3 *in vivo*, we transfected an expression vector for the haemagglutinin HA1 epitope-tagged hSur-2 (hSur-2e) into adenovirus-transformed 293 cells¹⁰ expressing both 289- and 243-residue E1A proteins which, respectively, contain and lack CR3 (ref. 11). The large E1A protein was

specifically co-precipitated when the extract was immunoprecipitated with anti-HA1 monoclonal antibody (Fig. 1d).

Western blotting of the AB, CB and DB protein fractions (Fig. 1a) with anti-hSur-2 antibody revealed approximately ten times more hSur-2 in the CB than in the DB fraction (data not shown). This had been obscured in our initial analysis of proteins bound to GST–E1A under non-stringent conditions by additional CB-fraction polypeptides binding to both wild-type and mutant columns. Gel filtration revealed that most of the hSur-2 in the D fraction eluted as expected for a globular, monomeric protein of M_r ~150K (data not shown). However, the hSur-2 in the CB fraction eluted ahead of a 670K marker, at M_r ~1,000K–2,000K (Fig. 2a). Cdk8 and cyclin C, human homologues of yeast *Saccharomyces cerevisiae* holoenzyme⁴ subunits Srb-10 and Srb-11 (refs 12, 13), are also present in a complex of high relative M_r ¹⁴ that binds to the E1A CR3 and herpes simplex virus VP16 activation domains¹⁵. Cdk8 and cyclin C co-eluted with hSur-2 from the gel-filtration column (Fig. 2a), well ahead of RNA polymerase II (Pol II) and general transcription factors (GTFs) TFIIE, F and H in the CB fraction.

To determine whether hSur-2, Cdk8 and cyclin C reside in a single complex analogous to yeast Mediator^{4,5}, peak gel-filtration-column fractions containing these proteins (Fig. 2a) were purified on a HiTrap Q anion-exchange column (Fig. 1a). All three proteins, and a newly described mammalian homologue of yeast Med7 (ref. 16), co-purified in the Q fraction (Figs 1a, 2b, lane 6). Cdk8, cyclin C, hSur-2 and hMed7 were co-immunoprecipitated with anti-Cdk8 antibody (Fig. 2b, lane 1), demonstrating that they exist in a single complex of high M_r . The complex bound to GST–E1A and GST–VP16, but not to GST–E1A H160Y (Fig. 2b, lanes 3–5). The complex eluted from GST–VP16 (but not GST–E1A) in 0.3 M KCl. Comparison of silver-stained polypeptides in the Q fraction and the GST–VP16 eluate (Fig. 2b, lanes 7 and 8) revealed that virtually all of the ~30 polypeptides in the Q fraction bound to the GST–VP16 column. These polypeptides also co-eluted during salt gradient elution of an analytical Mono-Q column (data not shown). Hence the Q fraction contained a highly purified multisubunit complex of ~30 polypeptides that bound to the activation domains of E1A and VP16.

We assayed the transcription activity of this proposed human homologue of the yeast Mediator in transcription reactions, using partially purified GTFs from which the complex had been separated by gel filtration (CBS fraction; Figs 1a, 2a, c). Gal4–E1A activated transcription from the $G_5\Delta$ MPL (Ad2 major late promoter with five upstream Gal4 sites) template containing activator-binding-sites in reactions with recombinant TFIIB and protein fractions AB, CB and DB (Fig. 3a, lane 1). Although the concentration of Pol II and TFIIE, -F and -H were comparable in the CB and CBS fractions (Fig. 2c), activated transcription was substantially reduced when the CBS fraction replaced the CB fraction (Fig. 3a, lane 2). Activation was fully restored when Q fraction was added (Fig. 3a, lanes 3 and 4). The Q fraction did not stimulate basal transcription from the $G_5\Delta$ MPL template. Results were similar with Gal4–VP16 (Fig. 3b). The Q fraction did not contain the previously described complex human RNA Pol II holoenzyme¹⁷, because it did not contain detectable Pol II, TFIIE or -F (Fig. 2c) and did not support transcription in the absence of Pol II and GTFs supplied by the CBS fraction (Fig. 3b, lanes 11 and 12). Based on this transcriptional activity and its homology with at least three yeast Mediator subunits, we refer to the Q-fraction multiprotein-complex as hMediator.

To confirm that the hMediator complex contained the transcription-stimulating activity in the Q fraction, we tested whether a GST–E1A column that binds hMediator (Fig. 2b) could deplete transcriptional activity in protein fractions from early in the purification. Pooled C and D fractions were passed twice over a GST–E1A column, or a control GST–E1A H160Y column, at a salt concentration that prevents TFIID binding⁷. Flow-through from the wild-type column contained about one-third as much of the hMediator subunits assayed (hSur-2, Cdk8 and cyclin C) as flow-

[†] Present address: Department of Molecular Medicine, Institute of Biotechnology, University of Texas Health Sciences Center, San Antonio, Texas 78245, USA.

through from the control H160Y column (Fig. 3c). The GTFs assayed were not depleted by the wild-type column (Fig. 3c). Activated, but not basal, transcriptional activity was reduced in reactions using the wild-type compared to the control column flow-through (Fig. 3c: compare lanes 1 and 5 with lanes 2 and 6). High levels of activated transcription were restored when the highly purified hMediator complex in the Q fraction was added (Fig. 3c, lanes 3, 4, 7 and 8), demonstrating that hMediator represents the activity required for activated transcription which was depleted by the wild-type E1A column. In a highly purified transcription system, the Q fraction inhibited activated transcription (Fig. 3d). Thus, the partially purified system contained a factor(s), in addition to GTFs and upstream stimulatory activity¹⁸, which was required for the stimulation of activated transcription by hMediator.

We analysed the *in vivo* function of hSur-2 in transient transfection assays. Increasing amounts of an hSur-2 expression vector inhibited Gal4-E1A activation, but not activation by Gal4-VP16

(Fig. 4a). Figure 4b shows our interpretation of the inhibitory effect of hSur-2 overexpression on Gal4-E1A activation. Western blotting (not shown) indicated that hSur-2 expression at the highest concentration of expression vector was at least tenfold higher than endogenous hSur-2. We propose that, under these conditions, excess hSur-2 interacts with the Gal4-E1A CR3 zinc-finger, precluding interaction with the hMediator complex through its hSur-2 subunit. As hSur-2 overexpression did not inhibit Gal4-VP16 activation, we propose that the VP16 activation domain interacts with a subunit of the hMediator other than hSur-2. Consistent with this model, there does not appear to be a *sur-2* homologue in the yeast genome, and E1A CR3 does not activate in yeast (T.G.B. and A.J.B., unpublished observations), whereas VP16 does.

We also analysed the effect of hSur-2 overexpression on activation by mammalian activators. Elk1 is an ETS-family transcription factor activated by a mitogen-stimulated Ras-MAP kinase pathway¹⁹. Overexpression of hSur-2 inhibited activation by

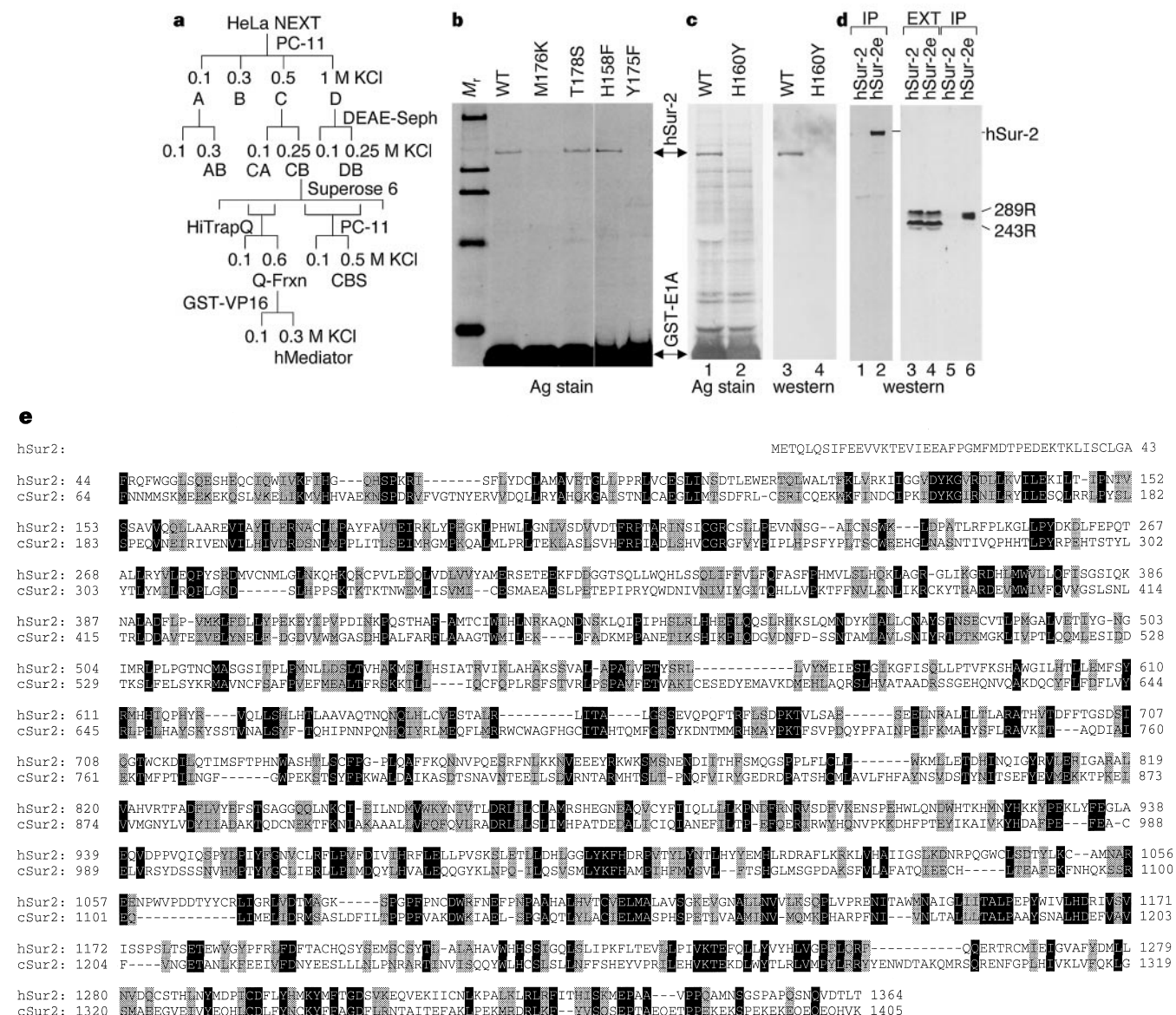


Figure 1 Identification of hSur-2. **a**, Fractionation of HeLa nuclear extract (NEXT). **b**, GST-E1A CR3 (wild type, WT) or indicated mutants bound to beads were incubated with D-fraction, washed and eluted with glutathione. Bound proteins were visualized by SDS-PAGE and silver staining. **c**, As in **b**, except eluates were analysed by silver staining and western blotting with antisera against an EST-derived 88-residue peptide. **d**, Extracts from 293 cells transfected with expression

vectors for HA1-tagged hSur-2e or non-tagged hSur-2 were immunoprecipitated with anti HA-1 monoclonal antibody. Total extract (EXT; lanes 3 and 4) or immunoprecipitates (IP; lanes 1, 2, 5 and 6) were subjected to western blotting with anti-hSur-2 (lanes 1 and 2) or anti-E1A (lanes 3-6) antibody. **e**, Alignment of hSur-2 and the homologous region of the 1,586-amino-acid *C. elegans* sur-2. Identical residues are highlighted in black and similar residues in grey.

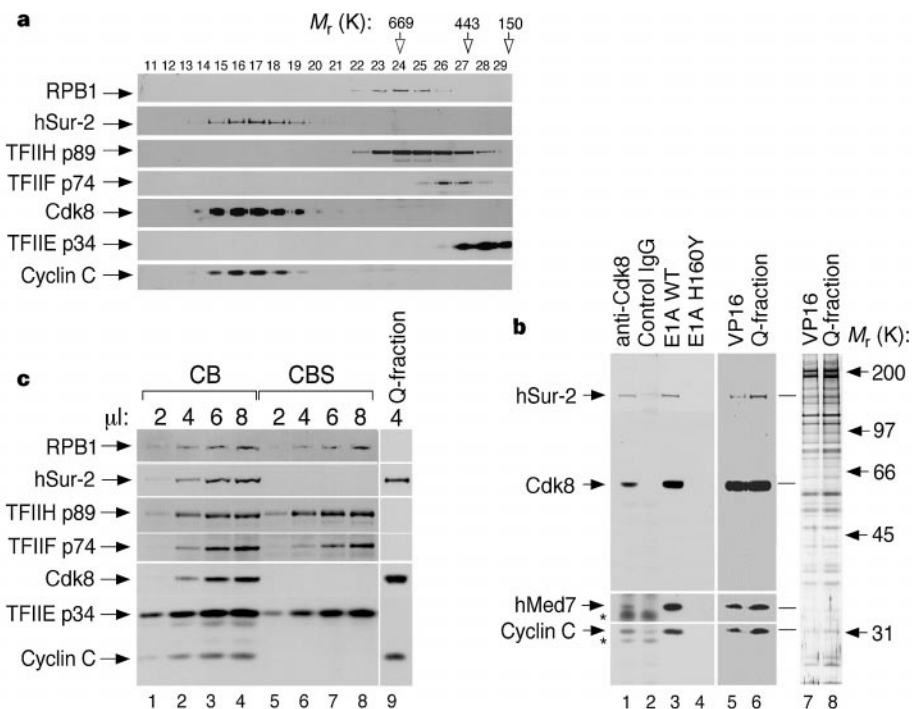


Figure 2 Purification and characterization of hMediator. **a**, Western blots of Superose 6 gel-filtration column fractions. Open arrows indicate marker protein peaks. **b**, Lanes 1–6, western blots of Q fraction (lane 6); Q fraction immunoprecipitated with anti-Cdk8 (lane 1) or control rabbit antibody (lane 2); Q

fraction bound to GST-E1A WT (lane 3), GST-E1A H160Y (lane 4) and GST-VP16 (lane 5). Asterisk, background from antibody. Lanes 7, 8, silver-stained gel of Q fraction bound to GST-VP16 (lane 7), and of total Q fraction (lane 8). **c**, Western blot of CBS, CB and Q fractions.

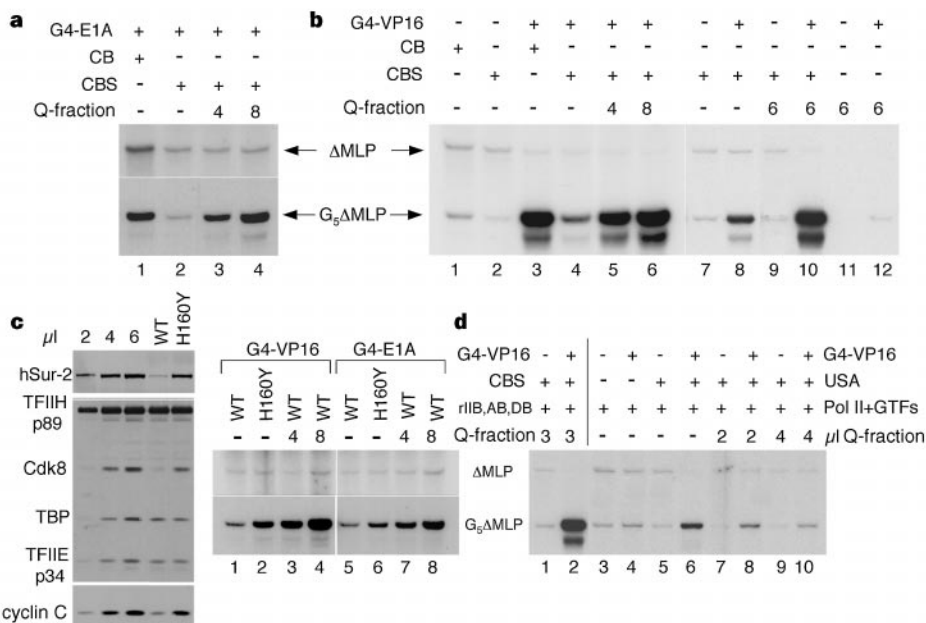


Figure 3 hMediator transcriptional activity. **a**, *In vitro* reactions using templates with ($G_5\Delta$ MLP) and without (Δ MLP) Gal4-binding sites, rTFIIB and fractions AB and DB. Gal4-E1A, fractions CB, CBS, and μ l of Q fraction added as indicated. Top portion was exposed for four times longer than the bottom. **b**, Reactions as in **a** with Gal4-VP16 in place of Gal4-E1A. Lanes 1–6 and 7–12 were from separate experiments. **c**, Left, western blot of fraction C–D after two passages over a GST-E1A WT or GST-E1A H160Y column (2, 4 and 6 μ l of

C + D were analysed as standards); right, transcription reactions using the C + D fraction passed over GST-E1A WT or GST-E1A H160Y, plus rTFIIB and fraction AB plus Gal4-VP16 (lanes 1–4) or Gal4-E1A (lanes 5–8). Q fraction was added as indicated. **d**, Reactions with rTFIIB and fractions AB, DB and CBS as indicated (lanes 1 and 2) and with rTFIIB, rTFIIE, rTFIIF, plus purified TFIIF and Pol II (Pol II + GTFs) (lanes 3–10). Gal4-VP16, USA and μ l Q fraction added are indicated.

Gal4–Elk1 in response to activated MEK kinase (Fig. 4c). In contrast, Gal4–ATF-2, which is regulated by a stress-activated MAP-kinase cascade²¹, was only slightly inhibited (Fig. 4c). These results indicate that Elk-1 targets hSur-2, whereas ATF-2 does not.

Alternative mammalian multiprotein complexes containing

homologues of yeast Srb and Mediator proteins have been characterized that have fewer polypeptides than the hMediator complex described here^{16,22–25}. The ~30-subunit complex we characterized stimulates activation by at least two classes of activation domains, represented by E1A CR3 and VP16. It is unclear whether this large

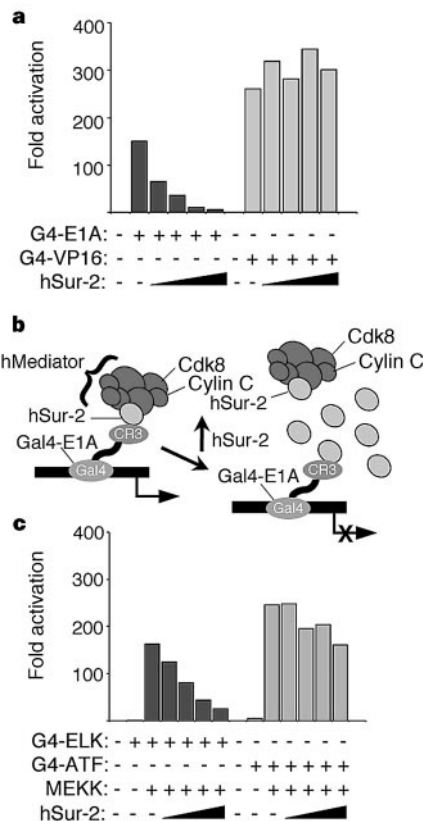


Figure 4 Overexpression of hSur-2 inhibits activation by specific activation domains. **a**, A luciferase reporter gene driven by a TATA box with four upstream Gal4 sites was transfected into HeLa cells either with (+) or without (-) expression vectors for Gal4-E1A or Gal4-VP16 and hSur-2 (0.1, 0.2, 0.4 and 0.8 μ g), as indicated. Ratio of luciferase activity in the presence and absence of activator is plotted. **b**, Model for inhibition of Gal4-E1A activation by overexpression of hSur-2. Excess hSur-2 competes with endogenous hMediator for binding to the E1A CR3 zinc-finger. **c**, Transfection experiment as in **a**, except with expression vector for Gal4-Elk1, Gal4-ATF2 and constitutively activated MEKK.

hMediator complex dissociates into simpler subcomplexes when exposed to high salt concentration or alternative chromatographic matrices, or whether mammalian Srb and Mediator proteins are distributed between several distinct and stable multi-subunit co-activator complexes. Genome-wide expression analysis in yeast indicates that the Srb4 subunit of the yeast Mediator is generally required for transcription by Pol II, whereas other Mediator subunits regulate subsets of yeast genes²⁶. Genetic analysis of *C. elegans* sur-2 (ref. 6) and our results on Elk-1 (Fig. 4c) indicate that the sur-2 subunit of metazoan Mediator complexes is a target of activation domains regulated by Ras-MAP kinase signal-transduction cascades that are important for intercellular communication. □

Methods

Affinity columns. GST-E1A (121–223) beads (30 μ l) incubated for 75 min at 25 °C with 3 mg D-fraction in 0.7 M KCl (Fig. 1b, c) or 400 μ g Q-fraction in 0.6 M KCl (Fig. 2b) were washed twice with 0.5 M KCl and twice with 0.1 M KCl, all in D buffer⁷, and eluted with 10 mM glutathione, 50 mM Tris at pH 8.0. GST-VP16 (413–490) beads (0.4 ml) were incubated with 2 mg Q fraction in 0.1 M KCl (Fig. 2b), washed with 0.1 M KCl and eluted with 0.3 M KCl, all in D buffer.

hSur-2 cDNA cloning. A 27-amino-acid sequenced peptide from 6 pmol hSur-2 (ref. 9) matched a single dbEST entry (accession no. Z24977) which was used to screen a cDNA library²⁷ for isolation of two partial cDNA clones that were ligated into a full-length 4.4-kb cDNA through the single EcoRI site 332 bp 5' to the polyA tail.

hMediator purification. 75 mg CB fraction (Fig. 1a) derived from 2.6 g HeLa nuclear extract⁷ was subjected to Superose 6 gel-filtration in 0.25 M KCl D buffer (12 runs on a 16 $\times$ 500 mm column). The hSur-2 peaks detected by western blotting (7 mg) were dialysed to 0.1 M KCl D buffer, applied to a 1-ml HiTrap Q-column (Pharmacia), and eluted with 0.6 M KCl D buffer. Peak protein fractions were pooled and dialysed to 0.1 M KCl D buffer (Q fraction; 5.5 mg).

In vitro transcription. This was as described⁷ with 60 ng pG5 Δ MLP(G-) (five Gal4 sites upstream of the Ad2 major late promoter (MLP)) and 60 ng p Δ MLP(G-) (MLP alone) upstream of 200 and 400-bp G-less cassettes, respectively. Reactions with the purified system used 160 ng rTFIIA, 40 ng rTFIIB, 80 ng rTFIIE, 150 ng rTFIIF, 2 μ l rTFIID, 1 μ l TFIIF purified from the CB fraction by chromatography on hydroxypatite and phenyl Sepharose, and 6 μ l Pol II purified by immunoaffinity chromatography with monoclonal antibody 8WG16.

Transient transfection assays. Superfect reagent (Qiagen) with 60% confluent cells was used with 10 μ g hSur-2 in pCS2+ per 10-cm dish (Fig. 1d). For Fig. 4, we used 0.1 μ g Gal4-E1A (ref. 28), Gal4-VP16 (ref. 28), pFA-Elk1 (Stratagene), pFA-ATF2 (Stratagene) and 0.2 μ g pDA-MEKK (ref. 20) expression vectors, as indicated, 2 μ g pGAL4-M2-Luc reporter plasmid (ref. 28), and 0.3 μ g pRL-TK (Promega) expressing Renilla luciferase per 6-cm dish. Firefly luciferase activity was normalized to Renilla luciferase activity.

Received 11 January; accepted 15 March 1999.

- Whyte, P., Williamson, N. M. & Harlow, E. Cellular targets for transformation by the adenovirus E1A. *Cell* **56**, 67–75 (1989).
- Zhu, L. et al. Growth suppression by members of the retinoblastoma protein family. *Cold Spring Harbor Symp. Quant. Biol.* **59**, 75–84 (1994).
- Arany, Z., Sellers, W. R., Livingston, D. M. & Eckner, R. E1A-associated p300 and CREB-associated CBP belong to a conserved family of coactivators. *Cell* **77**, 799–800 (1994).
- Meyer, V. E. & Young, R. A. RNA polymerase II holoenzymes and subcomplexes. *J. Biol. Chem.* **273**, 27757–27760 (1998).
- Myers, L. C. et al. The Med proteins of yeast and their function through the RNA polymerase II carboxy-terminal domain. *Genes Dev.* **12**, 45–54 (1998).
- Singh, N. & Han, M. sur-2, a novel gene, functions late in the let-60 ras-mediated signaling pathway during *Caenorhabditis elegans* vulval induction. *Genes Dev.* **9**, 2251–2265 (1995).
- Boyer, T. G. & Berk, A. J. Functional interaction of adenovirus E1A with holo-TFIID. *Genes Dev.* **7**, 1810–1823 (1993).
- Webster, L. C. & Ricciardi, R. P. trans-Dominant mutants of E1A provide genetic evidence that the zinc finger of the transactivating domain binds a transcription factor. *Mol. Cell. Biol.* **11**, 4287–4296 (1991).
- Brauer, A. W., Oman, C. L. & Margolies, M. N. Use of o-phthalaldehyde to reduce background during automated Edman degradation. *Analyt. Biochem.* **137**, 134–142 (1984).
- Graham, F. L., Smiley, J., Russell, W. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36**, 59–74 (1977).
- Flint, J. & Shenk, T. Adenovirus E1A protein: paradigm viral transactivator. *Annu. Rev. Genet.* **23**, 141–161 (1989).
- Tassan, J.-P., Jaquenoud, M., Leopold, P., Schultz, S. J. & Nigg, E. Identification of human cyclin-dependent kinase 8, a putative protein kinase partner for cyclin C. *Proc. Natl. Acad. Sci. USA* **92**, 8871–8875 (1995).
- Liao, S. M. et al. A kinase-cyclin pair in the RNA polymerase II holoenzyme. *Nature* **374**, 193–196 (1995).
- Rickert, P., Seghezzi, W., Shanahan, F., Cho, H. & Lees, E. Cyclin C/CDK8 is a novel CTD kinase associated with RNA polymerase II. *Oncogene* **12**, 2631–2640 (1996).
- Gold, M. O., Tassan, J.-P., Nigg, E. A., Rice, A. P. & Herrmann, C. H. Viral transactivators E1A and VP16 interact with a large complex that is associated with CTD kinase activity and contains CDK8. *Nucleic Acids Res.* **24**, 3771–3777 (1996).
- Jiang, Y. W. et al. Mammalian mediator of transcriptional regulation and its possible role as an end-point of signal transduction pathways. *Proc. Natl. Acad. Sci. USA* **95**, 8538–8543 (1998).
- Greenblatt, J. RNA polymerase II holoenzyme and transcriptional regulation. *Curr. Opin. Cell Biol.* **9**, 310–319 (1997).
- Kaiser, K. & Meisterernst, M. The human general co-factors. *Trends Biochem. Sci.* **9**, 342–345 (1996).
- Janknecht, R. & Nordheim, A. Gene regulation by Ets proteins. *Biochem. Biophys. Acta* **1155**, 346–356 (1993).
- Faris, M., Kokot, N., Lee, L. & Nel, A. E. Regulation of interleukin-2 transcription by inducible stable expression of dominant negative and dominant active mitogen-activated protein kinase kinase in Jurkat T cells. Evidence for the importance of Ras in a pathway that is controlled by dual receptor stimulation. *J. Biol. Chem.* **271**, 27366–27373 (1996).
- Denhardt, D. T. Signal-transducing protein phosphorylation cascades mediated by Ras/Rho protein in the mammalian cell: the potential for multiplex signalling. *Biochem. J.* **318**, 729–747 (1996).
- Sun, X. et al. NAT, a human complex containing Srb polypeptides that functions as a negative regulator of activated transcription. *Mol. Cell* **2**, 213–222 (1998).
- Cho, H. et al. A human RNA polymerase II complex containing factors that modify chromatin structure. *Mol. Cell. Biol.* **18**, 5355–5363 (1998).
- Gu, W. et al. A novel human SRB/MED-containing cofactor complex, SMCC, involved in transcription regulation. *Mol. Cell* **3**, 97–108 (1999).
- Ryu, S., Zhou, S., Ladurner, A. G. & Tijan, R. The transcriptional cofactor complex CRSP is required for activity of the enhancer-binding protein Sp1. *Nature* **397**, 446–450 (1999).
- Holstege, F. C. P. et al. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
- Mes-Masson, A. M., McLaughlin, J., Daley, G. Q., Paskind, M. & Witte, O. N. Overlapping cDNA clones define the complete coding region for the P210c-abl gene product associated with chronic myelogenous leukemia cells containing the Philadelphia chromosome. *Proc. Natl. Acad. Sci. USA* **83**, 9768–9772 (1986).
- Bryant, G. O., Martel, L. S., Burley, S. K. & Berk, A. J. Radical mutagenesis reveals TATA-box binding protein surfaces required for activated transcription *in vivo*. *Genes Dev.* **10**, 2491–2504 (1996).

Acknowledgements. We thank R. Kornberg for anti-mMed7 antibody and P. Rickert for anti-CDK8 and anti-cyclin C antibodies; M. Carey, W. Huang, S. Rundlett, S. Smale, J. Stevens and P. R. Yew for advice and comments; and C. Eng for technical assistance. This work was supported by the NIH. T.G.B. was initially supported by a postdoctoral fellowship from the California Division of the American Cancer Society.

Correspondence and requests for materials should be addressed to A.J.B.

correction

Elasticity and rheology of iron above 220 GPa and the nature of the Earth's inner core

Ho-kwang Mao, Jinfu Shu, Guoyin Shen, Russell J. Hemley, Baosheng Li & Anil K. Singh

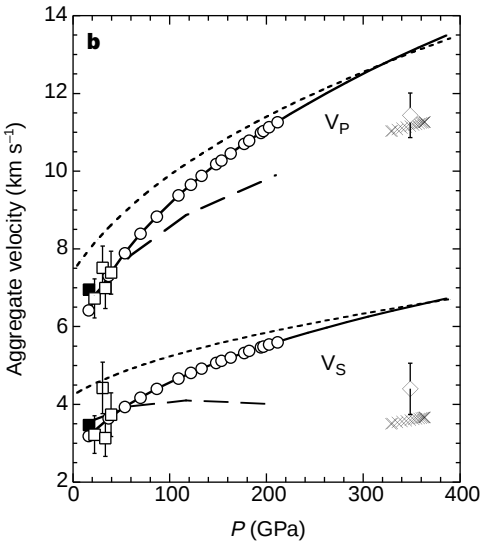
Nature 396, 741–743 (1998)

We regret that some numbers from an earlier iteration of our calculations were mistakenly used in Fig. 2b and Table 1. The correct figure and table are presented here.

In Fig. 2b, the aggregate V_P of Fe calculated from diamond-cell data of K and G (Fig. 2a) are in excellent agreement with low-pressure ultrasonic data, and compare favourably with high-pressure shock-wave and inner-core values at high temperatures ($dV_P/dT = -2.9 \times 10^{-4} \text{ km s}^{-1} \text{ K}^{-1}$), thus strengthening our original conclusions.

The single-crystal V_P anisotropies for Fe at 39 GPa and 211 GPa (in our original Fig. 4), as calculated from C_{ij} values based on isostress assumption, are now similar in magnitude ($\sim 8\%$) as well as in orientation (minima at a - and c -axes and maximum at 45° to the c -axis). Other parts of this Letter, including the discussion, remain unchanged.

We thank I. Jackson, G. Steinle-Neumann, S. Singh and L. Dubrovinsky for pointing out the inconsistency. □



Comparison of results from experiments and theories with seismic observations. Symbols are defined in the original figure.

Table 1 Elasticity of h.c.p. Fe at 298 K and high pressures

<i>P</i> (GPa)	Density (g cm ⁻³)	<i>C</i> ₁₁ (GPa)	<i>C</i> ₁₂ (GPa)	<i>C</i> ₃₃ (GPa)	<i>C</i> ₁₃ (GPa)	<i>C</i> ₄₄ (GPa)	<i>K</i> (GPa)	<i>G</i> (GPa)	<i>v</i> _P (km s ⁻¹)	<i>v</i> _S (km s ⁻¹)	Run no.
16.5	9.00*						297	108	6.95	3.47	3
39	9.67*	500	275	491	284	235	351*	134	7.40	3.73	1
39	10.09	747	301	802	297	215	455	224	8.64	4.72	†
211	12.61†	1533	846	1544	835	583	1071†	396	11.26	5.61	2
211	12.80	1697	809	1799	757	421	1085	445	11.45	5.90	†

The ABC of DNA

Amplification, blood DNA, cloning kits, detectors and other aids to handling or preparing DNA make up this week's selection of recently introduced products.

Quantum Prep

From Bio-Rad

www.bio-rad.com

Multiple-sample DNA preparation in 96-well format for robotic sequencing

The latest addition to the Quantum Prep range from Bio-Rad is a 96-well-format miniprep kit for the preparation of multiple samples of plasmid DNA for high-throughput automated sequencing. Yields are up to 15 µg plasmid DNA per well, so concentrations are sufficient for immediate use in sequencing reactions. This minimizes the number of preps required and can eliminate DNA desalting and concentration steps.

Reader Service No. 100



By-pass the salt, with Bio-Rad's new kit.

for cDNA library screening, primer synthesis and nucleic-acid hybridization. The UV crosslinker provides a short-wave radiation exposure chamber for bonding nucleic acid to a medium, usually before hybridization.

Reader Service No. 102

DNAetox

From Sterogene Bioseparations

www.sterogene.com

Remove endotoxin from DNA preparations

DNAetox is a solid-phase reagent for the removal of endotoxin from plasmid and synthetic DNA intended for gene therapy applications. Binding capacity is up to 1,000 EU per ml. Affinity of the immobilized ligand for DNA is low, for a claimed mass yield that typically exceeds 90%. DNAetox can be regenerated with 1 M NaOH and sterilized by autoclaving. The reagent is used either in a batch or in a chromatography column.

Reader Service No. 101

HybriLinker HL-2000

From UVP

www.uvp.com

Hybridizer and crosslinker combined

The HL-2000 HybriLinker hybridization oven combines the HB-1000 hybridizer with a 254-nm ultraviolet crosslinker in a single unit. The hybridizer creates the conditions



Cooking up hybrids: UVP's combination oven.

Thermo Sequenase II

From Amersham Pharmacia

www.apbiotech.com

A DNA sequencing enzyme evolves

Thermo Sequenase II is the successor to Thermo Sequenase, which was launched in July 1995. Thermo Sequenase II has several advantages over its predecessor, including: shorter cycling times; greater efficiency in difficult-to-read regions of DNA (trouble spots with high G+C content); tolerance to high-salt environments; and better incorporation of dyes for easier clean-up steps.

Reader Service No. 103

DNA purification system

From AutoGen

www.autogen.com

Sequence-quality DNA

The 850 Alpha DNA purification system purifies sequence-quality DNA from BACs, plasmids and M13, and has new protocols for plant material and mouse tail. When the run is completed the DNA is ready to use in applications such as fluorescent sequencing, PCR and restriction digests.

Reader Service No. 104

Universal support

From Beckman Coulter, Inc.

www.beckmancoulter.com

Universal support for DNA or RNA synthesis

Beckman Coulter's new universal linker CPG (controlled pore glass) solid-support system uses a volatile, non-salt cleavage and deprotection reagent. The manufacturers recommend the support for high-throughput production facilities, where the fact that it does not require a secondary purification

step to remove salt comes into its own.

Reader Service No. 105

Clone 18S

From Kamiya

www.kamiyabiomedical.com

Anti-DNA polymerase β monoclonal antibody for DNA-repair research

This monoclonal (clone 18S) is directed against DNA polymerase β, a 39K enzyme which fills single-nucleotide gaps produced by the base-excision repair pathway of mammalian cells. The antibody binds DNA polymerase β from human, cow, mouse, rat, hamster and *Xenopus*. Crossreactivity with other species has not been tested. The antibody does not inhibit polymerase activity

Reader Service No. 106

FastTag

From Vector Laboratories www.vectorlabs.com

A fluorescent nucleic-acid labelling kit

The FastTag Texas Red nucleic-acid labelling kit couples non-radioisotopic labels to DNA, RNA or oligonucleotides. Heat- or photo-activation is used to uniformly incorporate a universal FastTag linker arm into the nucleic acid. The fluorescent Texas Red maleimide label is covalently attached to the nucleic acid via the FastTag linker arm. Applications for these labelled nucleic acids include *in situ* hybridization, Southern and northern blot hybridization, PCR amplification, and colony and plaque hybridization.

Reader Service No. 107

Biotech Photometer

From WPA

www.wpald.co.uk

A photometer for DNA analysis

The WPA UV1101 Biotech Photometer is configured to simplify routine analysis of DNA/RNA, including $A_{260,280}$ purity values. There is no warm-up time and the unit will accept samples down to 70 µl. The UV1101



Not a UV/VIS spectrophotometer, from WPA.

is marketed as compact, low-cost alternative to the UV/VIS spectrophotometer and is also cheaper to run. The photometer can be used in both the UV- and visible regions and comes with filters at 260, 280 and 600 nm, enabling bacterial growth rates to be measured. Other filter wavelengths are available.
Reader Service No. 108

Molecular biochemicals

From Roche www.roche.com/diagnostics
Call 01273 480 444 (UK) for a copy

The 1999 Molecular Biochemicals Catalogue from Roche Diagnostics covers over 2,500 products and includes sections on apoptosis, non-radioactive systems and quantitative PCR. Under the PCR heading there is an expanded range of reagents for nucleic-acid studies, as well as the LightCycler system for performing PCR in less than 30 minutes.
Reader Service No. 109

mtDNA Extractor WB Kit

From Wako www.wakousa.com
Nal procedure for whole blood

Wako's mtDNA Extractor WB Kit provides an efficient protocol for isolation of mtDNA from human whole blood, regardless of the presence of anticoagulants. The procedure

can be performed in about 90 minutes. The mtDNA preparation has low contamination of genomic DNA and is suitable for use with restriction enzymes and PCR amplification.
Reader Service No. 110

DNAzol and DNAzol BD

From Molecular Research Centre
www.mrcgene.com

New reagents for DNA isolation from blood
DNAzol and DNAzol BD (for 'Blood Derivatives') are ready-to-use reagents that extract genomic and/or viral DNA from whole blood. The reagents contain a novel guanidine-detergent lysing mixture that promotes selective precipitation of DNA. Small aliquots of whole blood (200 µl to 1 ml) provide sufficient DNA to visualize stained samples in agarose gels following electrophoretic separation and for successful amplification reactions. Genomic DNA yield from human whole blood is 30–40 µg DNA per ml blood.
Reader Service No. 111

RT-PCR kit

From Hybaid www.hybaid
Single-tube PCR

The ExpeRT RT-PCR kit has been developed for RT-PCR in a one-tube format. A new

buffer system is used to achieve RT-PCR of fragments up to 6 kb in a single step. The buffer composition is also claimed to reduce problems with secondary structures in RNA templates and to increase the specificity of primers. The kit comprises AMV reverse transcriptase which can be used at temperatures as high as 60°C and Hybaid's ProofSprinter (Taq/Pwo) enzyme mix. The kit is suitable for *in situ* RT-PCR as well as for tube-based RT-PCR.
Reader Service No. 112

TdT monoclonal

From BioGenex www.biogenex
Antibody to terminal deoxynucleotidyl transferase

The monoclonal TdT88 reacts specifically with TdT in formalin-fixed, paraffin-embedded tissue sections. In normal tissue, TdT is found in a subpopulation of cortical thymocytes and bone-marrow lymphocytes. Abnormal TdT expression has been reported in acute lymphocytic leukaemia and chronic myeloid leukaemia. The antibody is available in concentrated form or optimized for Super-Sensitive detection systems. The antibody is currently available for *in vitro* diagnostic use.
Reader Service No. 113

GenePrint PowerPlex

From Promega UK www.euro.promega.com
Single-tube amplification and sequencing

The GenePrint PowerPlex 1.2 System is optimized for use with the ABI PRISM 310 genetic analyser and PRISM 377 DNA sequencer. It facilitates simultaneous single-tube amplification and two-colour detection of eight polymorphic STR loci plus Amelogenin. High-throughput analysis is achieved by comparing amplified sample DNA fragments directly with the allelic ladder present for each locus.
Reader Service No. 114

TOPO cloning kits

From Invitrogen www.invitrogen.com
New vectors for cloning and sequencing

Two new TOPO cloning kits are available that allow ligation of PCR products generated with proofreading or nonproofreading thermostable polymerases. The vectors are provided linearized and activated with topoisomerase I, which permits 5-minute ligation. The TOPO TA Cloning Kit for Sequencing and the Zero Blunt TOPO PCR Cloning Kit for Sequencing are engineered to contain the T7, T3, M13 forward and M13 reverse priming sites as close to the insert as 33 base pairs. This minimizes the amount of plasmid DNA sequence that needs to be read before reading the sequence of the clone.
Reader Service No. 115

ADVERTISEMENTS

contact: lightcycler@TIB-MOLBIOL.de

TIB
MOLBIOL

HybProbes
for the **LightCycler**

- Mutation Detection
- Quantitative PCR

Custom Synthesis of Hybridization Probes
for the new LightCycler under licence of Roche Diagnostics

READER ENQUIRY NO. 7

The Mouse KitTM
for Gene Targeting in ES Cells

- ✓ Complete with embryonic stem (ES) cells, feeder cells and essential reagents (including FBS) for engineering one knockout mouse strain
- ✓ Low-passage AB2.2-Prime ES Cells[®] (129/SvEv)
- ✓ Uniform, high-quality ESQ feeder cells (mitotically inactivated STO cells) engineered for *neo* and *puro* resistance
- ✓ Prequalification of ESQ feeder cells and reagents for use with AB2.2-Prime ES Cells

*Manufactured by Lexicon Genetics Incorporated

Fax: (619) 535-0071

STRATAGENE

Tel: (619) 535-5400 or (800) 424-5444
www.stratagene.com

READER ENQUIRY NO. 6

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT

PROTEIN SEQUENCING
Membrane-bound, solid or solution samples

£12 per cycle
5 cycles minimum - no set up charge

For more information please contact Dr. Joe Gray:
Phone/Fax: +44 191 222 8612
Email: JOE.GRAY@newcastle.ac.uk

DNA sequencing (ABI 377-96KL, 877 Catalyst)
also available at competitive prices

READER ENQUIRY NO. 2

Diagnostic Reagents

Pefachrome[®]-Series: chromogenic and fluorigenic substrates

Pefabloc[®]-Series: serine proteinase inhibitors

Purified snake venom proteins
(e.g. Protac[®], Ecarin, Batroxobin)

PENTAPHARM LTD
CH 4002 Basel Fax: +41 61 319 96 19
<http://www.pentapharm.ch>

READER ENQUIRY NO. 16

Neutra-Plasma 50

From Neutra-Plasma

Decontamination? Pop it in the microwave

The Neutra-Plasma 50 beaker, launched this week at the Laboratory Show in London's Olympia, decontaminates reaction tubes for PCR in only 30 seconds using a domestic microwave oven. The lidded container comprises an enclosed quartz vessel which produces UV light at the germicidal wavelength of 253.7 nm. This introduces pyrimidine dimers into contaminating DNA molecules, rendering them incapable of amplification in the PCR process. The beaker may also be of use in other areas where microbial contamination is a problem.

Reader Service No. 116



Oven-to-bench: the Neutra-Plasma 50.

and size of the 8 linear PMT channels match those of the plates, making direct measurements of optical signals possible.

Reader Service No. 117

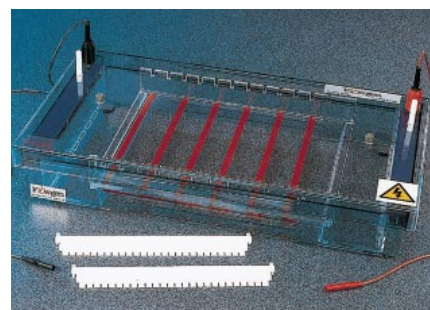
Multichannel detector

From Hamamatsu

www.hamamatsu.com

A detector for DNA sequencing

The R5319 Multichannel PMT detects low-light-level luminescence from samples in microtitre plates with single-photon counting on each channel. Microtitre plates are widely used for DNA sequencing. The pitch



Flowgen's new tank: gets on with MADGE.

Flowgen is recommended for multi-sample analytical PCR as it allows fast and convenient screening. It is fully compatible with 8- and 12-well multichannel pipettes and special sample combs have been designed for this purpose. A 10 × 25-cm unit is also available, retaining full multichannel pipette compatibility with lower consumable use. When using a 96-well PCR machine, all samples can be screened on a single gel. This system is also compatible with MADGE gels.

Reader Service No. 118

Gel tank

From Flowgen

www.flowgen.co.uk

A gel apparatus for screening PCR products

The new 30 × 25-cm horizontal tank from

EZSwitch

From Robbins Scientific

www.robsci.com

Gel loaders for the ABI 377 DNA sequencer

EZSwitch DNA sequence gel loaders are compatible with the ABI 377 DNA sequencer with 96-lane upgrade. The gel loader is a lightweight, hand-held dispenser equipped with eight 5-μl syringes. The EZSwitch gel loader changes spacing by moving all of the syringes in the array, not by bending the syringe needles, which can make loading samples difficult. Syringe needles also have flexible tips that resist damage and can be easily replaced.

Reader Service No. 119

GMS microarray analyser

From Genetic Microsystems

www.geneticmicro.com

A DNA microarray system brought down to size

The GMS Microarray Analysis System, say Genetic Microsystems, allows individual researchers to prepare reproducible microarrays of DNA and rapidly capture high-resolution images of hybridized arrays, in their own labs, in experiments of their own design. The system consists of two bench-top modules: the GMS 417 Arrayer and the GMS 418 Array Scanner. The computer-controlled system uses a Windows interface and includes software tools for instrument operation and data analysis.

Reader Service No. 120

These notes are compiled in the Nature office from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

ADVERTISEMENTS

Pepsyn Ltd

High Quality Peptide Synthesis

'Standard' and 'non-standard'
Fast delivery
Competitive prices
Also advisory services

Contact: Dr J A Smith, Chairman, Pepsyn Ltd
School of Biological Sciences, Life Science Building
The University of Liverpool, Liverpool L69 7ZB, UK
Tel: 0151 794 4329 Fax: 0151 794 4349
e-mail: J.A.Smith@liverpool.ac.uk
Website: www.pepsyn.co.uk

READER ENQUIRY NO. 9

Lysyl Endopeptidase

(*Achromobacter* Protease 1)

Specifically cleaves all Lys-X
bonds including Lys-Pro.
Highly Stable, Water Soluble



Wako BioProducts
Wako Chemicals USA, Inc.
1600 Bellwood Road
Richmond, VA 23237
Telephone: (804) 271-7677
Facsimile: (804) 271-7791
(800) 992-9256
bioproducts@wakousa.com

READER ENQUIRY NO. 31

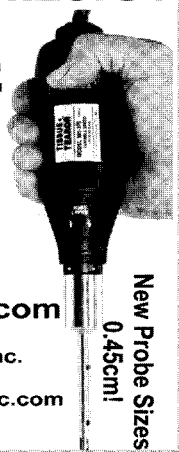
Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

Check
Us
Out

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com



New Probe Sizes
0.45cm

READER ENQUIRY NO. 54

Check out
our homepage at
<http://www.resgen.com>

BACs

YACs

PACs

cDNAs

Research Genetics provides clones and colony membranes for a number of human, mouse, and rat libraries as well as a custom screening service for the libraries.

Research Genetics, Inc.
U.S. or Canada 800-533-4363
U.K. 0-800-89-1393
FAX 256-536-9016

READER ENQUIRY NO. 30